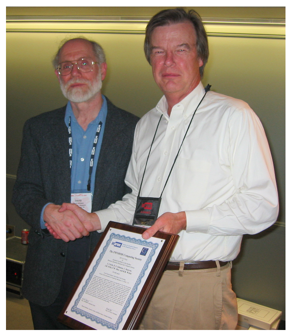
ICS NEWS
INFORMS Computing Society Newsletter

Inside This Issue

2007 ICS Prize Goes to Janos Csirik, David Johnson, Claire Kenyan, James Orlin, Peter Shoer, and Richard Weber

Michael Ball, University of Maryland
mball@rhsmith.umd.edu



Mike Ball (right) presents the ICS Prize Certificate to David Johnson (left).

The 2007 ICS prize was awarded to Janos Csirik, University of Sieged, David Johnson, AT&T Labs, Claire Kenyan, Brown University, James Orlin, MIT, Peter Shoer, MIT, and Richard Weber, University of Cambridge, for the results in their article, "On the sum of squares algorithm for bin packing," *Journal of the Association of Computing Machinery* 53:1 (2006), 1–65.

The paper presents a very comprehensive analysis of an algorithm for the on-line bin packing problem. The algorithm, which is deceptively simple, is known as the "sum of squares algorithm." The authors show that under a very wide class of input distributions the algorithm is asymptotically optimal. Further, the analysis leads to simple algorithmic modifications that lead to asymptotic optimality for input distribution classes where the basic algorithm is not asymptotically optimal.

ICS Prize continued on page 26 ▷

2007 ICS Student Paper Award Goes to Amit Partani and David Morton

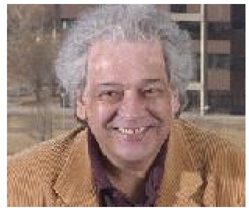
Jonathan Eckstein, Rutgers University
jeckstei@rutcor.rutgers.edu

The 2007 INFORMS Computing Society Student Paper Prize is awarded to Amit Partani and his advisor David Morton of the University of Texas at Austin, for the paper "Adaptive jackknife estimators for stochastic programming." The paper addresses a fundamental question in stochastic programming: efficiently using sampling to estimate the quality of solutions to problems whose uncertain parameters take an infinite or very large number of values. The committee found the paper to be well written, combining high-level understanding of multiple ar-

Jonathan Eckstein (right) presents the Student Paper Award Certificate to Amit Partani (left)

ICS Student Paper Award continued on page 26 ▷

*Visit **http://list.informs.org/pipermail/computing-society/** for the ICS monthly enews archive.*



Message from the Editor

Harvey J. Greenberg
University of Colorado at Denver
harvey.greenberg@cudenver.edu

Congratulations to the ICS Prize and Student Award winners! The competition continues to make the selections difficult.

Our Chair's message highlights our progress during the short time he has been Chair. Robin Lougee-Heimer continues to be our liaison with COIN-OR and reports what is happening. She is also the incoming ICS Chair, and as Chair-Elect, Robin organized the clusters for the INFORMS Seattle meeting — a record 81 sessions! Her report follows that of the ICS participation at INFORMS Puerto Rico. The ICS Biennial Meeting will be in Charleston, 2009.

Get involved: Volunteer your help.

ICS now has an Education Committee, and a brief report is given here by its Chair, Jill Hardin. Their primary task is to develop curriculum guidelines for OR/CS education, and Allen Holder chaired a panel discussion in Seattle to solicit feedback on the 13-page report. Jill gives a succinct summary of the report and the Seattle feedback, and she solicits further discussions. Thanks to Bill Hart, we have a blog to post your suggestions and opinions. We have reports from our other project leaders: Al Holder, Editor of the *Mathematical Programming Glossary* and Rob Dell, Director of the *Leading Edge Tutorials*.

I am pleased to present a profile of Karla Hoffman, a longtime leader of ICS and its ancestor organizations. Karla has a strong background in all aspects of our profession: research, teaching, and problem-solving. She chooses problems whose solutions benefit people.

We begin a new column with this issue, called "Dear Dr. ORCS." This brainchild of John Chinneck has the form of a lovelorn column.

The *INFORMS Journal on Computing* has a new editor: John Chinneck. He has jumped into this, devoting much of his time to making the online system work. We thank David Kelton who not only served the previous two terms as Editor, but also took over when last year's Editor was unable to continue. John launched a new Area: *Computational Biology and Medical Applications*.

I am pleased to publish three technical articles by excellent researchers in our field. Anna Nagurney describes "supernetworks," which has blossomed at The University of Massachusetts at Amherst under her leadership. Andrew Kusiak tells us about data mining, which has been a focus of his research for decades, and more recent connections with "innovation science." Finally, Hemant Bhargava and Juan Feng describe how auctions apply to "search-engine advertising," one of the components for e-commerce.

Sadly, I include an obituary of Lloyd Clarke, who met an untimely death on September 26.

I added *News from Related Communities*, and a *Humor* column. I expect both of these to be a regular part of *ICS News*.

This year's issues were produced with LaTeX. The source files are posted, which anyone can use as a template. Comments welcome.

Enroll your students in ICS FREE through December 20!
http://computing.society.informs.org/freepromo.php

Message from the Chair

John W. Chinneck
Carleton University
chinneck@sce.carleton.ca

My term as ICS Chair will come to a close at the end of 2007, so this is my last *Message from the Chair* column for *ICS News*. What a great two years it's been! Together we've started a number of terrific projects. It's mind-boggling to look at the list: established a new website, formalized a link to COIN-OR, provided a home for the *Mathematical Programming Glossary*, established the ICS History Archive, started a monthly news email, started the ICS Leading Edge Tutorial series, awarded the first ICS Student Paper Awards, established the new Harvey J. Greenberg Award for Service to ICS, held a great meeting in Coral Gables, ran a wildly successful student sign-up program, established an Education Committee, added some features to the newsletter, ran record-breaking numbers of sessions at the INFORMS meetings and, well, I'm sure I've probably forgotten a few others. None of this could have been accomplished without you, the members. It's your enthusiasm and involvement that makes a job like this so much fun. Most of the time I felt like an orchestra conductor: all I did was wave a little white baton, but it was you folks who were actually making the music.

Best of all, I'm sure that the pace will continue under incoming ICS Chair, Robin Lougee-Heimer. This is the person who organized an incredible 81 ICS sessions for the Seattle INFORMS meeting, handily surpassing the previous record. I know she has some ideas for new projects already coming down the pipeline: look out! . . . But this is also a great opportunity for you to get involved, meet some new people, and generally combine your professional and academic interests with a little fun.

Thanks to each and every one of you for your involvement and enthusiasm. As they say in the *Hitchhiker's Guide to the Galaxy*, "So long, and thanks for all the fish!"

COIN-OR

Robin Lougee-Heimer, IBM
robinlh@us.ibm.com

The ***CO**mputational **IN**frastructure for **O**perations **R**esearch* (COIN-OR) is the premier website devoted to open-source software for the operations research community. Hosted by INFORMS, COIN-OR is home to over two dozen projects and a burgeoning community. COIN-OR encourages new project contributions by providing extensive tools and infrastructure for collaborative project development. For more information about COIN-OR's history, goals and projects, see "COIN-OR Pays Off" ▹ [*OR/MS Today*, October 2005], This article highlights 2007 new projects and new users. Also, please visit http://www.coin-or.org, or look under "Projects" on the ICS website.

New Projects

Five new projects debuted on COIN-OR in 2007. In addition to the GAMSlinks project [*ICS News*, Spring 2007], the following four new projects have begun.

COIN-OR Graph Classes by Philip Walton (Junction Solutions, Inc.), is a collection of network representations and algorithms to facilitate the development and implementation of network algorithms. A few graph representations and many algorithms are supplied as part of the library.

LaGO by Stefan Vigerske and Ivo Nowak (Humboldt-University of Berlin), is the Lagrangian Global Optimizer software package for the global optimization of nonconvex mixed-integer nonlinear programs. LaGO comes with interfaces to GAMS and AMPL.

OBOE by Jean-Philippe Vial, Alain Haurie, and Nidhi Sawhney (University of Geneva), is the Oracle Based Optimization Engine, an implementation of the Analytic Center Cutting Plane Algorithm whose goal is to solve convex but non-differentiable problems using only subgradient information. OBOE has been developed over the last fifteen years at the University of Geneva.

OS by Robert Fourer, Jun Ma (Northwestern University), and Kipp Martin (University of Chicago), is the Optimization Services project. OS is designed to provide a set of standards for representing optimization instances, results, solver options, and communication between clients and solvers in a distributed environment using Web Services.

(continued on page 26)

Education Committee

Jill Hardin, Virginia Commonwealth University
jrhardin@vcu.edu

ICS formed an Education Committee at the 2006 INFORMS meeting (in Pittsburgh). The Committee members are:

- Jill Hardin (Chair), Virginia Commonwealth University, jrhardin@vcu.edu
- Kevin Furman, Exxon-Mobil, kevin.c.furman@exxonmobil.com
- Allen Holder, Trinity University, aholder@trinity.edu
- David Rader, Rose-Hulman Insitute of Technology, David.Rader@rose-hulman.edu
- Cesar Rego, University of Mississippi, crego@bus.olemiss.edu
- Advisors: Harvey J. Greenberg and Ariela Sofer.

The Committee's charge was to outline appropriate curricula for undergraduate students planning to pursue work at the OR/CS interface. Given the diversity of programs offering courses in operations research, and given the diversity of curricula across institutions in general, the Committee determined that the best approach is to outline a list of skills, rather than a list of courses, that would prepare students for work at the OR/CS interface. A 13-page report is posted at http://computing.society.informs.org/eduCom.php.

(continued on page 27)

Mathematical Programming Glossary

Allen Holder, Trinity University
aholder@trinity.edu

The ICS *Mathematical Programming Glossary* [http://computing.society.informs.org/glossary.php] continues to broaden its coverage to better suit the optimization community. In addition to normal maintenance and slight corrections, the following are either complete or will be in the near future:

- A tour of stochastic optimization, including a wider collection of terms;
- An overhaul of the *Glossary*'s coverage of complexity, with a robust collection of modern terms and a tour to help guide students through this material;
- A supplement on reformulation techniques for integer programs;
- Addition of terms associated with mathematical programs with equilibrium constraints (MPECs).

Some confusion about the URL followed the transition from Harvey Greenberg's personal web page to the INFORMS server, but this has dissipated, and the *Glossary* currently receives about 43,000 hits per week.

You can assist the *Glossary* by using it as an educational resource, suggesting new terms, or authoring a tour and/or a supplement. Also, if you use the *Glossary* as an author or for course notes, please cite the *Glossary*, as given in the *General Information* section.

Leading Edge Tutorials

Rob Dell, Naval Postgraduate School
dell@nps.navy.mil

The informal motto of ICS is "INFORMS' Leading Edge for Computation and Technology." True to this spirit, ICS has started tutorials at INFORMS conferences on topics of interest to ICS members.

Dr. Vernon Austel, IBM T. J. Watson Research Center, presented *Perl in an Hour* at the INFORMS international meeting in Puerto Rico.

The Seattle INFORMS meeting had three Leading Edge Tutorials. Professor Dave Alderson, Naval Postgraduate School, presented *Critical Thinking for Complex Network Systems: Trends, Tools, and Techniques for the OR Analyst*, and Professor Pascal Van Hentenryck, Brown University, presented *(Almost) Transparent Parallel and Distributed Optimization*. (See the abstracts ▷.) The third was a part of the main *TutORials* program: Professor Leon Lasdon, University of Texas at Austin, and Dr. Janos Pintér, President of Pintér consulting, jointly presented *Computational Global Optimization*, which was published in the INFORMS *TutORials in Operations Research* Series (see http://tutorials.pubs.informs.org/).

We are already thinking about the next INFORMS meeting in Washington, DC. Please visit http://computing.society.informs.org/lEdge.php to learn how you can contribute.

Free Student Membership

For a limited time, ICS offers free Student Membership. If you are an academic (or can validate a student status), you can sign up your students at no cost. This offer will expire December 20. Visit http://computing.society.informs.org/freepromo.php and simply enter your name, email, and institution, followed by the students you wish to sign up. There is a limit of 5 students per faculty. The students will receive an email welcoming them as ICS Student Members, indicating that you signed them up. They will be asked to go to another INFORMS site to complete the process by providing information about themselves. This is an opportunity that you do not want to miss. Your students will receive all benefits from ICS membership, and they represent our future.

Since this promotion started September 20, 33 faculty from 26 institutions have signed up 97 students. There is still time to sign up more students and encourage them to get involved with ICS.

ICS Web Site

Pascal Van Hentenryck, Brown University, pvh@cs.brown.edu,
Laurent Michel, University of Connecticut, ldm@engr.uconn.edu

In the last year, we have completely redesigned the ICS website and expanded its contents in many directions. Some of the recent additions include:

- Summary of ICS Participation at Seattle INFORMS meeting, by Robin Lougee-Heimer
- Link to archive of the Chair's monthly enews (maintained by INFORMS)
- ICS Blog, by Bill Hart
- Student Community page — see *Top 10 Ways for Students to Get Involved in ICS*

Check it out at http://computing.society.informs.org/ for recent news and up-to-date information on ICS activities. You will find exciting projects, such as the *MP Glossary*, the COIN-OR partnership, Leading Edge Tutorials, and our Education Committee. You will also find community news, information about ICS meetings, *ICS News* (our newsletter), and the *INFORMS Journal on Computing*.

INFORMS Puerto Rico

Jun Ma and Bob Fourer (Northwestern University) organized 14 sessions for the INFORMS International Meeting held in Puerto Rico, July 8–11, 2007. You can still see the ICS cluster online ▹. You can also see pictures that Harlan Crowder took of ICS folks at the Reception by visiting http://www.harlanpics.com/public/puertorico07/.

INFORMS Seattle — Record-breaking ICS Activity

Robin Lougee-Heimer, IBM
robinlh@us.ibm.com

The INFORMS 2007 Seattle meeting was a phenomenal display of ICS activity including technical sessions galore, an anniversary celebration, tutorials, prize winners, progress on all ICS project fronts, a new hazing ritual (!), and a blog to capture it all.

Technical Program

The old record of 46 ICS-sponsored sessions established in Pittsburgh (2006) was shattered by a phenomenal 81-session lineup in Seattle. The program consisted of (i) our traditional cluster, (ii) special topics clusters, (iii) clusters co-sponsored with other subdivisions and Invited cluster organizers, and (iv) jointly-sponsored sessions in a non-joint clusters. The 81 sessions appeared with ICS attribution in the program, but are not in one list on the conference website. The ICS special topics clusters were: *Bioinformatics and Systems Biology* (5 sessions), *COIN-OR* (6 sessions), *Constraint Programming and Optimization* (5 sessions), and *Metaheuristics-Sponsored* (2 sessions). The joint clusters were *Computational Methods for Data Mining* with the Data Mining Section (7 sessions), *Computational Optimization and Software* with the Optimization Society (9 sessions), and *Metaheuristics* with an Invited cluster chair (3 sessions). The joint sessions in non-joint clusters were collaborations with Artificial Intelligence (1 session), Applied Probability Society (3 sessions), CPMS, the Practice Section (1 session), Health Applications Section (4 sessions), Information Systems Society (4 sessions), INFORM-ED (3 sessions), Optimization Society (1 session), and Telecommunications (4 sessions). The traditional cluster (23 sessions) included the ICS 2006 Prize session, ICS 2007 Prize session, ICS 2007 Student Paper Award session, ICS Leading Edge Tutorial session, and the ICS 10th Anniversary and Quiz Show session.

Thanks to everyone who presented in and organized ICS-sponsored sessions, and a special thank you to the cluster organizers:

- Ionut Aron, IBM,
- Brian Borchers, New Mexico Institute of Mining and Technology,
- Victoria Chen,University of Texas at Arlington,
- Allen Holder, Trinity University,
- Gary Kochenberger, University of Colorado at Denver,
- Sanjay Mehrotra Northwestern University,
- Leo Lopes, University of Arizona,
- Robin Lougee-Heimer, IBM,
- Nick Sahinidis, Carnegie Mellon University,
- David Woodruff, University of California.

A personal highlight of the program was the Anniversary Quiz Show organized by ICS Board Member Steve Dirkse (GAMS). The formidable intellect of ICS members unbridled in game-show repartee was an absolute hoot. And complete with catering, too — thanks to GAMS. See Steve's report ▹, which follows, for more on this commemorative event.

Business Meeting

The Anniversary Quiz Show segued into the business meeting conducted by Chair John Chinneck (Carleton University) and attended by 65 people, according to the official roster. Reports from the officers, updates from our project leaders, and new business filled the agenda. Characters from *The Simpsons* filled the .ppt deck, keeping the fun-factor high throughout the perfectly-paced presentation.

"We're still rich!" reported Secretary/Treasurer Jeff Linderoth (University of Wisconsin - Madison). Jeff estimates our year end balance will be in excess of $30K and reported membership to be solid at 489 (up from 423 last year). A good part of the increase is due to new student members. Special thanks to Harvey Greenberg (University of Colorado at Denver), Tod Morrison (University of Colorado at Denver), Laurent Michel (University of Connecticut), and INFORMS staff for the student sponsorship sign-up initiative. Remember — Students Join Free Until December!

Reports given by project leaders appear in this newsletter, as well an update on the *INFORMS Journal on Computing* from Editor John Chinneck.

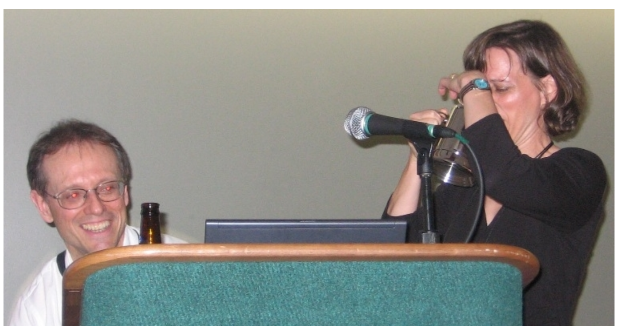
New business included the changing of the guard. Thanks to following officers whose terms were ending:

- Lou Hafer (Simon Fraser University) and Nick Sahinidis for serving as Board Members,

- Jeff Linderoth for serving as Secretary/Treasurer,
- Robin Lougee-Heimer for serving as Vice-Chair (taking over as Chair),
- John Chinneck for serving as Chair.

And thanks to continuing Board Members Rob Dell (Naval Postgraduate School), Pascal Van Hentenryck (Brown University), Steve Dirkse (GAMS), and Matt Saltzman (Clemson University); and, to Webmasters Pascal Van Hentenryk and Laurent Michel .

In stepping down as Chair, John Chinneck assumes his new role as the first "ICS Historian." John got right to work making history by instituting a new tradition. Popping opening a beer, John produced a gleaming silver mug engraved with "ICS" and "The Big Cheese." Pouring the beer into the stein, he explained the Canadian basis of the ritual. John ceremoniously passed the title of Chair to the Chair-elect in passing the Cup for an inaugural sip. And so, the new ("pop & pass" | "chug & chair"?) tradition was born. Thanks to John for his exemplary leadership, innovation, and camaraderie.



John passes the *ICS Stein* to Robin, who shows her dislike for beer.

And thanks to Bill Hart (Sandia National Laboratories), you can read more about the Seattle meeting and add your own comments about the conference on our blog at http://computing.society.informs.org/serendipity. Check it out. Next up — the INFORMS 2008 Annual Meeting in Washington, DC!

INFORMS Seattle — 10th Anniversary Celebration & Quiz Show

Steve Dirkse, GAMS
steve@gams.com

What do you get when you combine creative ICS Presidents who don't take no for an answer, a full-loving panel of ICS experts, a hilarious scorekeeper, and a session chair willing to try something different? We found out in Seattle at the *ICS 10th Anniversary Celebration and Quiz Show*, where our panel of eight contestants battled for the prize: first-run autographed ICS protectors.

Team GAMS:	Team MPL:
Jerry Brown	Dick Barr
Harlan Crowder	Bjarni Kristjansson
Alex Meeraus	Ariela Sofer
John Tomlin	Bill Stewart

The teams first played a round of "ICS Jeopardy," resulting in negative scores for both teams! Things improved, though, when we switched to "Society Feud." Next it was on to "Survivor: Seattle," where each team would vote somebody off their side. One person on each side could win immunity from this in the Slide Rule Challenge. Jerry Brown and Ariela Sofer each won immunity, although rumor has it Ariela went high-tech with an HP-35. The teams each staged a coup, voting off Alex and Bjarni. The reorganization was effective as the teams really hit their stride in the "Who wants to be an Optimizaire" segment, strategically using their lifelines (the audience was ***very*** helpful) to rack up big points.

John Chinneck shows the pocket protector he made as a prize for the winners (but he gave it to everyone).

The audience was involved right from the start — several of the panelists had prepared quiz questions with which they periodically peppered the crowd. The savvy ICS fans usually had no problems answering, although they were almost stumped by some of Jerry's artifacts. The final segment, "Whose Face Is That Anyway," was played by one and all. In the end, scorekeeper Jeff Linderoth announced that Team GAMS had won a narrow victory, which correlated highly with GAMS Development's sponsorship of the wine and cheese for the session.

Ariela Sofer especially enjoys the prize while her teammates look on: Alex Meeraus, Dick Barr, and Bill Stewart.

In a meeting with so many great ICS-sponsored sessions, this one was uniquely special. This was made possible by the good humor of the participants and the audience, my quiz show team of John Chinneck, Jeff Linderoth, and Robin Lougee-Heimer, and the decades of contributions by fun-loving members. See all of the Show content at http://www.gams.com/ics/quizShowSeattle/, and visit the ICS site, as more will be posted.

ICS Biennial Meeting

Matt Saltzman
mjs@clemson.edu

Our Biennial Meeting will be held January 11-13, 2009, at the Francis Marion Hotel in Charleston, South Carolina. This year's theme is *Operations Research and Cyberinfrastructure*. Please send your ideas to any member of our Committee co-Chairs:

- John Chinneck, chinneck@sce.carleton.ca
- Bjarni Kristjansson, bjarni@maximal-usa.com
- Matt Saltzman, mjs@clemson.edu
- Chris Starr, starrc@cofc.edu

We would particularly like to hear from you if you are interested in organizing an invited session. Visit our website at http://www.ics2009.org/ to learn more about our meeting and how you can participate.

Call for Photos

Do you fancy yourself a photographer? Do you attend ICS events with a digital camera? *ICS News* needs photos of ICS people at events. Even if your photo is not used in *ICS News*, it may appear in a forthcoming photo gallery at our web site. If you have photos (or questions), please send to the *ICS News* Editor. ▷

Member Profile: Karla L. Hoffman

In Passaic, NJ, a town adjacent to Clifton, Abe and Bertha Rakoff gave birth to their second daughter, Karla, who now lives with her husband Allan in Clifton, VA, a town named after the one in New Jersey. This is a sign that she would never veer from her roots.

At the age of 13, while working for her father's auto parts store, Karla already showed OR acumen with these suggestions:

Demand-driven location. *When I started working in the store, all parts were organized by type of part and then by car make and year. I noticed that there were a few parts (e.g., batteries) that were retrieved far more often than most other parts and yet were often not easily reachable. I suggested placing these most widely wanted parts in a separate section where they could be more easily retrieved. That was my first O. R. success!*

Queuing reduction. *I also thought it better to have the customer pay for the part as soon as it was retrieved (if he/she so desired) . Originally, the parts department would retrieve the part and then the customer would pay for the part at the front of the store. I suggested having a cash register in the parts department, thereby allowing the customer to pay immediately. One additional cash register improved service significantly.*

To Karla, this was "Nothing deep, just obvious ways of making operations better." To her father, these were great suggestions, which he immediately implemented. This foreshadowed her developing into a sophisticated problem-solver, using OR from her education and CS from her first professional job.

Karla received her B.S. in Mathematics from Rutgers University with a minor in economics. She then elected to pursue an MBA at George Washington University, having no idea that there was a field called OR that combined mathematics, economics, and business. It was after taking a course in optimization from Gerald Bracken that she realized there is a field that allowed her "to use analytic and human-relationship skills to solve problems that are critical to industry and government." It was an exciting time at GWU because of the arrival of Jim Falk, Tony Fiacco, and Garth McCormick. She worked under Jim Falk and learned her first programming language, Fortran, in order to code one of Jim's algorithms. When finishing her doctorate, Garth urged Karla to apply to the National Bureau of Standards (NBS). They had a talented OR group led by Alan Goldman. That is where Karla "lucked into" a job that exposed her to a great variety of challenging problems, while working with an extraordinarily talented team who remain her friends to this day.

I asked Karla how she got into the CS interface. She said, "While at NIST (NBS at the time), I was given the assignment of creating an OR toolbox for the department. It was, at that time, impossible to determine which algorithms and software worked well for what problems. That assignment and my work for both the Internal Revenue Service and Department of Transportation, both requiring software development, reinforced my view that working on applied problems requires a thorough knowledge of computer science."

The years at NBS shaped Karla's OR/CS pathway. She was the first mathematician to receive the Bureau of Standards' Applied Research Award. That same year, 1984, she received the U.S. Department of Commerce Silver Medal for Meritorious Service for her work with Manfred Padberg: *Advancing the use of combinatorial optimization techniques by government agencies*.

Karla began her service to ICS with its ancestor, *ORSA Special Interest Group in CS*, and she served on the Committee to propose that this be promoted to the *CS Technical Section*. She was then elected to be Secretary/Treasurer in 1979 and Vice Chair in 1980, thus becoming its third Chair in 1981. In 1985 Karla served on the Committee to organize the first biennial Conference on OR/CS. She served on the first Editorial Board of the *ORSA Journal on Computing*, 1987–1992, and she continues to serve on other editorial boards.

Karla has routinely been involved in professional society activities. She served on the Executive Committee of the Mathematical Programming Society, and worked to establish the Beale/Orchard-Hays Prize for Excellence in Computational Mathematical Programming. Karla is the founding Editor of the *Newsletter of the Committee on Algorithms* of the Mathematical Programming Society. Karla was President of INFORMS in 1999 and has served as Vice President of Finance and Chair of the Investment Committee. She also served as Treasurer of ORSA, and was actively involved in the ORSA-TIMS merger. Karla was inducted as an INFORMS Fellow in 2003, and she received the INFORMS Kimball Medal in 2005.

Her experiences at NBS set a foundation for her career at George Mason University, which began in 1984. She brought her applications into the classroom, making OR exciting for her students. It was this kind of attention to teaching for which she received the George Mason University Distinguished Professor Award in 1989. Karla served as Department Chair 1996–2001, during which time the Department changed its name to incorporate Systems Engineering (SE). Since SE and OR have always shared a history of looking at problems from a global perspective, the merging of these two departments allowed natural synergism among its faculty to occur.

No longer the unsophisticated teen from Passaic, Karla raised the bar on the problems she tackles. From early projects on model and code evaluation, she redirected her attention to several challenging problem areas. Her work on efficient crew scheduling set Karla on a pathway to apply OR/CS to other problems in aviation and in real-time scheduling of buses and trucks. When considering how to reduce congestion at major U.S. airports, she examined the reasons that airlines were scheduling flights in excess of that which the runways could handle and realized that existing government policies encouraged such behavior. Her background in auction theory — based on her consulting to the Federal Communications Commission on the allocation of spectrum through combinatorial auctions — provided an alternative mechanism that could more fairly and efficiently allocate runway access. From Karla's perspective, working on the runway congestion problem is a nearly perfect problem: It requires the use of economics, game theory, simulation, optimization and high-speed computing and has the potential to make a difference in many people's lives.

From Karla's perspective, working on the runway congestion problem is a nearly perfect problem: It requires the use of economics, game theory, simulation, optimization, and high-speed computing and has the potential to make a difference in many people's lives.

To test if market-based approaches could solve the congestion problem, she worked with Mike Ball to assemble a team consisting of faculty from Berkeley, George Mason, Harvard, MIT, and the University of Maryland, under the auspices of The National Center for Excellence for Aviation Operations Research (NEXTOR). This team designed, and ran for major decision makers within the aviation industry and policy makers within the FAA and DOT, a collection of strategic games to determine the impact of changes to government policy. Sophisticated simulation and optimization models were used to predict the outcomes (e.g., cost, schedule adherence, congestion) from each scenario. Such games were new to the industry, the policy makers, and the researchers. They succeeded because the games were realistic, the players were the eventual decision-makers, and there were well-designed models running in real-time to provide quick feedback.

Outside of her professional world, Karla is currently on the Board of Directors of the Parkinsons' Foundation of the National Capital Region. From working with those suffering from this devastating disease, she has come to realize how health care issues impact not only the patient, but also their families and the community. Karla is passionate about using her OR/CS skills to make a difference.

For recreation, Karla power-walks with friends, attends many cultural events in the DC area, cooks, and takes dance lessons. She prefers to spend her downtime walking on a beach and reading.

In closing, I quote Karla [*Journal of Computational and Applied Mathematics* 124:1-2 (2000), 341–360], showing some of how CS bears on the fulfillment of the promise of OR — to solve problems:

> *The availability of reliable software, extremely fast and inexpensive hardware and high-level languages that make the modeling of complex problems much faster have led to a much greater demand for optimization tools.*

Dear Dr. ORCS

This is a Q&A column with questions people have about the OR/CS interface.

Dear Dr. ORCS: The predecessor to the INFORMS Computing Society is the Computer Science Technical Section. This constitutes a name change from "Computer Science" to "Computing." Is there a difference?

John Chinneck: *Our new name, the Composting Society, reflects our proud commitment to the environment by recycling kitchen scraps to — no, wait, that's for INFARMS, not INFORMS, sorry. As I was saying, the Computing Society still embodies bilateral applications between OR & CS. However, the "computing" label is broader, so we see ICS as a broad-based society for anyone with interests in OR & Computing, including all of the CS interfaces.*

Dear Dr. ORCS: I have a problem. I have not written simulation programs in about 20 years, and when I last took a course on simulation they taught us how to use a package called "Simple Simon." Can I get hold of Simple Simon now?

Abhijit Gosavi: *I don't think so. You see, Simon has abandoned his simplicity and has taken to professional modeling instead. You will be surprised to find how glamorous he has become. He is nowadays called "ProModel."*

Dear Dr. ORCS: Who initiated research on sparse matrices?
— Saul Gass

Harvey Greenberg: *The concept of a sparse matrix and its importance in computation was noted by Watkins* [11] *in 1915. The field of sparse matrix methods for solving systems of linear equations began with linear programming. Dantzig and Orchard-Hays* [1,5,6] *brought it to the computer in the 1950s as part of their research program at The RAND Corporation. Several numerical analysis groups within the CS community broadened the use of sparse matrices, notably at IBM, in the late 1960s and early 1970s. Willoughby led this effort and produced the first Sparse Matrix Proceedings* [12]*. Tewerson's survey* [10] *captures essentials from the 1960s, but the vast developments were just on the verge of changing how we generate and solve large, sparse systems. The first Ph. D. theses were outside LP, by Rose* [8]*, George* [3]*, and Duff* [2]*. McNamee* [4] *produced the first sparse matrix package in 1971. The next two decades saw an explosive growth in theory and applications. Besides revisiting many algorithms to see how sparsity can improve performance, the graph-theoretic foundation emerged. Parter* [7] *used this in a very specialized way; Rose* [9] *broadened and deepened this foundation. Rose, George, and Duff have remained leading researchers in sparse matrices and associated algorithms for the past four decades and continue to publish new results. Large-scale mathematical programming continues to use sophisticated information structures, stemming from sparse matrix techniques and beyond.*

[1] G.B. Dantzig and W. Orchard-Hays, The Product Form of the Inverse in the Simplex Method, *Mathematics of Computation* 8:1 (1954), 64–67.

[2] I.S. Duff, *Analysis of Sparse Systems*, Ph. D. Thesis, Oxford University, UK, 1972. Note: this considers stiff ODEs.

[3] J.A. George, *Computer Implementation of the Finite Element Method*, Stanford University, Stanford, CA, 1971. Note: this considers ordering algorithms to avoid fill-in.

[4] J.M. McNamee, Algorithm 408: A Sparse Matrix Package (Part I), *Communications of the ACM* 14:4 (1971), 265–273.

[5] W. Orchard-Hays, Evolution of Linear Programming Computing Techniques, *Management Science* 4:2 (1958), 183–190.

[6] W. Orchard-Hays, *Advanced Linear Programming Computing Techniques*, McGraw-Hill, 1968.

[7] S. Parter, The Use of Linear Graphs in Gauss Elimination, *SIAM Review* 3:2 (1961), 119–130.

[8] D.J. Rose, *Symmetric Elimination on Sparse Positive Definite Systems and the Potential Flow Network Problem*, Ph. D. Thesis, Harvard University, Cambridge, MA, 1970. Note: this builds a graph-theoretic foundation.

[9] D.J. Rose, A Graph-Theoretic Study of the Numerical Solution of Sparse Positive Definite Systems of Linear Equations, *Graph Theory and Computing*, Academic Press, New York, NY, 183–217, 1972.

[10] R.P. Tewarson, Computations with Sparse Matrices, *SIAM Review* 12:4 (1970), 527–543.

[11] G.P. Watkins, Theory of Statistical Tabulation, *Publications of the American Statistical Association* 14:112 (1915), 742–757. (This became the *Journal of the American Statistical Association*.)

[12] R. Willoughby (Ed.), *Sparse Matrix Proceedings*, IBM Watson Research Center, 1969.

Dear Dr. ORCS: I carry out research in complexity theory and Operations Research. I prove theorems about NP-completeness, give polynomial time algorithms for OR problems, and occasionally give approximation algorithms for OR problems. Even though I often publish in CS journals, many colleagues in the OR/CS community don't seem to consider my research to be on the interface of OR and CS unless I also code algorithms and have empirical results. I think that this is prejudiced on their part, and I feel neglected. How should I respond?

— Lonely in Boston

Dear Lonely: *It's all a matter of what crowd you hang out with. Did you know that complexity theory got started in a bar? (Unfortunately, I don't know which one; finding it must have been intractable.) Maybe it would serve you well to note the quote used by Harvey Wagner* [Principles of Operations Research, 1969]*, "There is nothing so practical as good theory."* [Kurt Lewin, 1952].

Dear Dr. ORCS: I am new to the area of Operations Research, and I heard that I can win a million dollars if I can determine whether $P = NP$. As far as I can tell, $P = NP$ if, and only if, $P = 0$ or $N = 1$. How do I get my million dollars?

— New OR Researcher

Dear Newbie: *That particular solution was already given by Ted Swart at the Rutgers Advanced Research Institute in Discrete Applied Mathematics, 1987. (He went on to present a deeper answer. The shortcomings of his proof technique was analyzed by M. Yannakakis* [Expressing Combinatorial Optimization problems by Linear Programs, *Journal of Computer and System Sciences* 43:3 (1991), 441–466]*.)*

Dear Dr. ORCS: My boss wants a decision analysis system based on Bayesian belief nets to use for regular business decision-making. The end results will go to our executives, who are not OR/DA savvy. Should we use a general BBN application like Netica that makes the basis of decisions visible and maybe conveys the results in a written report? Or, should we build a custom application that hides the belief net from view but lets the executives play 'what if' for specific business questions?

— Confused

Dear Confused: *General vs. tailored is an age-old question when picking software. Tailoring it for your immediate decision support has advantages of discourse language and bringing out what is important to your executives. If your boss finds it useful, he will increase the scope and start to ask other questions that you had not anticipated. That's where a general* *system is advantageous. Hiding what's under the hood serves only to remove what you think would be daunting, keeping the interface extremely simple. I think you will find it profitable for the users to have the option of seeing what's driving the decision support system. Some may get into it and make better use of the BNN. There are a lot of BNN software systems out there, besides Netica...see* http://www.cs.ubc.ca/~murphyk/Bayes/bnsoft.html. *I know you have a schedule to meet, but if you have time, I suggest a careful assessment of off-the-shelf systems for your needs. As a stop-gap, you could code something quickly in Matlab — see Kevin Murphy's Bayes Net Toolbox at* http://bnt.sourceforge.net/.

Dear Dr. ORCS: While hosting a speaker in our university Operations Research/Management Sciences Speaker Series, I heard that the introductory course in OR that covers basic modeling and optimization was going to be taught by the CS faculty in one of the "big" schools in Cambridge, MA. With fewer and fewer departments in the US with "OR" in their names, I was wondering what you thought about computer scientists taking over the teaching of OR? Frankly, I was rather taken aback when I heard the news but I suspect that this may becoming more and more common. Interestingly, it is usually the "Artificial Intelligence" faculty that are considered the experts in optimization in CS departments.

Should we "rebrand" ourselves as CS folks? As the President of INFORMS recently said, "No one outside of OR has heard of us."

— Concerned

Dear Concerned: *One way to combat these 'takeovers' is to do it better and make the others know how you can offer their students a best education. Sometimes these takeovers begin out of frustration from an oblivious OR faculty. While this is not always the case, and may not be in what you've experienced, a concerned faculty should take such moves seriously in a positive sense — meet with the CS faculty and learn why they feel the need to teach their own OR course. The CS Dept probably has a shortage of faculty to teach the mainstream CS courses, so one can believe that they would happily give up a takeover under the right conditions. You can also think about having a core OR course that is taught on a rotating basis by OR and CS. Why not do the same for introductions to AI, Data Mining, and many other subjects in the interface. Their interest in teaching OR could be taken as a* ***good*** *sign of potential collaborations in teaching and research. Some of their students may switch to OR, or take it as a minor, due to the introduction by a CS faculty. Remember von Neumann's words, "In mathematics you don't understand things. You just get used to them." Some of us get too used to OR implanted in one department and may need to embrace the idea that a multidisciplinary approach to problem-solving must relate proactively with other disciplines.*

You raise another serious point: OR has had an identity crisis for decades. There are reasons for this, but let me just say that I spoke to this very point with respect to our relation with CS at the Coral Gables ICS Meeting (see my slides at http://computing.society.informs.org/activities.php*). Several approaches are being taken to address this. One is the formation of an Education Committee, chaired by Jill Hardin, who is in touch with ACM counterparts. A second is this newsletter and its broader distribution. A third is new outreach for JOC under John Chinneck's leadership. These, and other approaches, require some commitment by OR/CS leaders to make a difference.*

The Time is Now
The Leaders are You

If you have a question for Dr. ORCS, please send to the *ICS News* Editor. ▷ The questions can be serious or humorous (or both).

joc.pubs.informs.org

Message from the Editor of INFORMS Journal on Computing

John W. Chinneck ▷
editor_joc@mail.informs.org

I officially became the Editor-in-Chief of our affiliated journal, the *INFORMS Journal on Computing*, on July 1, 2007, so this is my first newsletter report in that capacity. Following in the footsteps of such notables as Harvey J. Greenberg, Bruce Golden, W. David Kelton, and Prakash Mirchandani, I've had big shoes to fill. Fortunately there is a collection of talented and dedicated people at the *JOC* who have been extremely helpful. My thanks go most especially to David Kelton, who stepped in again as Interim Editor-in-Chief when Prakash Mirchandani was sidelined by a family health crisis, while simultaneously acting as an Interim Area Editor, until I was able to take up the post. I have relied heavily on David for advice and help while finding my footing. In fact David actually assembled the summer 2007 edition of the Journal in which my first "From the Editor-in-Chief" column appeared.

It has been a busy couple of months. Several new editors have been recruited. ICS member Bob Fourer of Northwestern University has replaced me as the Area Editor for *Modeling: Methods and Analysis*. You may know Bob for his work on AMPL and NEOS. ICS member Winfried Grassmann of the University of Saskatchewan, a well-known researcher in the field, replaced David Kelton in his role as Interim Area Editor for *Computational Probability and Analysis*. In addition, the *Heuristic Search and Learning* area has grown steadily over the past several years, and was in dire need of additional Associate Editors, so two were appointed to help ease the load. They are Erwin Pesch of the University of Siegen in Germany, and ICS member Jean-Paul Watson of Sandia Labs.

(continued on page 27)

Supernetworks: The Origins, Some Applications, and Possibilities

Anna Nagurney ▷
nagurney@gbfin.umass.edu

Anna Nagurney is the John F. Smith Memorial Professor in the Department of Finance and Operations Management in the Isenberg School of Management at the University of Massachusetts at Amherst. She is the first female to be appointed to a named Professorship in the University of Massachusetts system. She is the Founding Director of the Virtual Center for Supernetworks and the Supernetworks Laboratory for Computation and Visualization at UMass Amherst. She received her AB, ScB, ScM, and PhD degrees from Brown University in Providence, Rhode Island. She has published more than 125 papers and eight books. Her most recent book is *Supply Chain Network Economics: Dynamics of Prices, Flows, and Profits* [Edward Elgar Publishing, 2006].

In an invited essay, *Navigating the Network Economy* [*OR/MS Today*, June 2000 ▷], I argued that we were in a new era of *Supernetworks*. Since that time the world has been transformed through events of historical proportions which have dramatically and vividly reinforced the dependence of our societies and economies on different networks from *physical* networks; i.e., transportation and logistical networks, communication networks, energy and power networks, to more *abstract* networks comprising: financial networks, environmental networks, social, and knowledge networks, and combinations thereof. Indeed, historical events over the past several years have graphically illustrated the interconnectedness, interdependence, and vulnerability of organizations, businesses, and other enterprises on network *infrastructure* systems. The decisions made by the users of the networks, in turn, affect not only the users themselves but others, as well, in terms of safety and security, profits and costs, timeliness of deliveries, the quality of the environment, etc. Hence, the understanding of the impacts of human decision-making on such networks is of paramount importance.

In this essay, I argue that Supernetworks are the *paradigm* for the modeling, analysis, and solution of complex problems in the information-based *Network Economy*. In particular, the supernetwork paradigm, as evidenced by my book [20], along with many articles and applications (see: http://supernet.som.umass.edu), is sufficiently general and yet elegantly compact to formalize the modeling and analysis associated with network systems on which humans interact. ***Super*** networks are

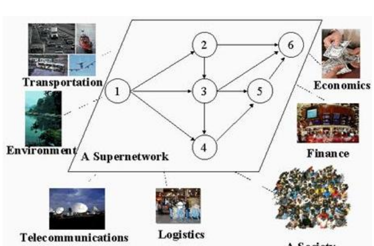


Figure SN-1. A supernetwork (▷ Larger picture)

networks that are **above and beyond** existing networks, which consist of nodes, links, and flows, with nodes corresponding to locations in space, links to connections in the form of roads, cables, etc., and flows to vehicles, data, etc. Supernetworks are conceptual in scope, graphical in perspective, and, with the accompanying theory, which is networked-based and predictive in nature.

The supernetwork framework, captures, in a unified fashion, decision-making facing a variety of decision-makers including consumers and producers as well as distinct intermediaries, such as financial brokers, electric power distributors, and electronic retailers. The decision-making process may entail weighting trade-offs associated with the use of transportation versus telecommunication networks. The behavior of the individual decision-makers is modeled as well as their interactions on the complex network systems with the goal of identifying the resulting flows and prices. By being able to predict the various flows based on network topologies and interactions amongst the decision-makers one gains deep insights into the vulnerabilities as well as the strengths of various linkages and network structures.

The Origins of Supernetworks

The origins of supernetworks can be traced to the study of transportation networks, telecommunication networks, and, interestingly, to biology, as reviewed in [20]. Below I highlight the origins of the term *supernetwork*.

In Transportation

In 1972, Dafermos [6] demonstrated, through a formal model, how a *multiclass* traffic network could be cast into a single-class traffic network through the construction of an expanded (and abstract) network consisting of as many copies of the original network as there were classes. She clearly identified the origin/destination pairs, demands, link costs, and flows on the abstract network. The applications of such networks she stated, "arise not only in street networks where vehicles of different types share the same roads (e.g., trucks and passenger cars) but also in other types of transportation networks (e. g., telephone networks)." Hence, she not only recognized that abstract networks could be used to handle multimodal transportation networks but also telecommunication networks! Moreover, she considered both user-optimizing and system-optimizing behavior, terms which she had coined with Sparrow [8] in 1969 (and which correspond, respectively, to Wardrop's (1952) first and second principles of travel behavior [37]). Her research was motivated, in part, by Beckmann, McGuire, and Winsten's 1956 book, *Studies in the Economics of Transportation* (see also [1]). In 1976, Dafermos [7] proposed an integrated traffic network equilibrium model in which one could visualize and formalize the entire transportation planning process (consisting of origin selection, or destination selection, or both, in addition to route selection, in an optimal fashion) as path choices over an appropriately constructed abstract network. The genesis and formal treatment of decisions more complex than route choices as path choices on abstract networks, that is, supernetworks, were hence reported as early as 1972 and 1976.

The importance and wider relevance of such abstract networks in decision-making, with a focus on transportation planning were accentuated through the term "hypernetwork" used by Sheffi [34], which he later [35] redefined as a "supernetwork." He recognized Dafermos' 1976 contributions and considered probabilistic-choice models. Thus, decision-making in a transportation context could be modeled as a "route" selection over an abstract network. The route, henceforth, referred to as a "path" to emphasize the generality of the concept, would correspond to a choice and the links to parts and pieces of the complete decision.

In Telecommunications

In his 1985 *American Scientist* article, Denning [9] discussed the internal structure of computer networks and emphasized how "protocol software can be built as a series of layers. Most of this structure is hidden from the users of the network." He then proceeded to ask the question, "What should the users see?" Denning answered the question in the context of the then National Science Foundation's Advanced Scientific Computing Initiative to make national supercomputer centers accessible to the entire scientific community. He said that such a system would be a network of networks, that is, a "supernetwork," and a powerful tool for science. Interestingly, he emphasized the importance of location-independent naming, so that if a physical location of a resource would change, none of the supporting programs or files would need to be edited or recompiled. Hence, in a sense, his view of supernetworks is in concert with that of ours in that nodes do not need to correspond to locations in space and may have an abstract association.

In 1979, Schubert, Goebel, and Cercone [33] had used the term in the context of knowledge representation as follows: "In the network approach to knowledge representation, concepts are represented as nodes in a network. Networks are compositional: a node in a network can be some other network, and the same subnetwork can be a subnetwork of several larger supernetworks. . . "

In 1997, the Illinois Bar Association considered the following to be an accepted definition of the Internet [11]: "the Internet is a supernetwork of computers that links together individual computers and computer networks located at academic, commercial, government and military sites worldwide, generally by ordinary local telephone lines and long-distance transmission facilities. Communications between computers or individual networks on the Internet are achieved through the use of standard, nonproprietary protocols." The reference to the Internet as a supernetwork was also made in The *Atlantic Monthly* in 1996 by Fallows [10], who noted that "The Internet is the supernetwork that links computer networks around the world."

Vinton G. Cerf, the co-developer of the computer networking protocol TCP/IP, in his keynote address to the Internet/Telecom 95 Conference [36], noted that at that time there were an estimated 23 million users of the Internet, and that vast quantities of the US Internet traffic "pass through internet MCI's backbone." He then went on to say, "Just a few months back, MCI rolled out a supernetwork for the National Science Foundation known as the very broadband network service or VNBS . . . VBNS is being used as an experimental platform for developing new national networking applications."

Decision-making on transportation and telecommunication networks can be done simultaneously through the supernetwork concept. For example, as demonstrated in [20], supply chain networks with electronic commerce, financial networks with intermediation, teleshopping versus shopping, telecommuting versus commuting, as well as transportation and location decisions in the Information Age formulated and solved within the supernetwork theoretical umbrella.

A variety of abstract networks in economics were studied in my 1999 book [16], which also contains extensive references to the subject. In [20] we have demonstrated that the abstract network concept also captures the interactions between/among the underlying networks of economies and societies. As noted in [17]: "The interactions among transportation networks, telecommunication networks, as well as financial networks is creating supernetworks. . . "

In Genetics

Interestingly, the term supernetworks has also been applied in biology, notably, in genetics. According to Noveen, Hartenstein, and Chuong [30], many interacting genes give rise to a gene network, with many interacting gene networks giving rise to a gene "supernetwork." They further state: "The function of a gene supernetwork is more complicated than a gene network. A gene supernetwork, for example, may be involved in determining the development of an entire limb while a gene network, working within the supernetwork, may be involved in setting up one of the axes of the limb bud." According to the same source, a gene supernetwork is defined as "a collection of gene networks which participate with each other during the morphogenesis of a specific structure, for example an organ, a segment, or an appendage." The authors then go on to discuss duplication, divergence, and conservation of a gene supernetwork and note that, as with gene networks, gene supernetworks can be duplicated during evolution, "thus giving rise to new structures which are the same as or different from the original structure."

Clearly, one of the principal facets of network systems today is the interaction among the networks themselves. For example, the increasing use of electronic commerce, especially in business to business transactions, is changing not only the utilization and structure of the underlying logistical networks but is also revolutionizing how business itself is transacted and the structure of firms and industries. Cellular phones are being used as vehicles move dynamically over transportation networks resulting in dynamic evolutions of the topologies themselves. Power outages in one part of the world may affect transportation and financial systems around the globe as the August 14, 2003 blackout demonstrated. The unifying concept of supernetworks with associated methodologies (optimization theory, network theory, variational inequalities, projected dynamical systems, etc.) allows one to explore the interactions among such networks as transportation networks, telecommunication networks, as well as financial networks, to capture the dynamic interactions and also to measure the associated risks and gains/losses.

Supernetworks and Applications

Supernetworks may be comprised of such networks as transportation, telecommunication, logistical and financial networks, among others. They may be multilevel as when they formalize the study of supply chain networks or multitiered as in the case of financial networks with intermediation. Furthermore, decision-makers on supernetworks may be faced with multiple criteria and, hence, the study of supernetworks also includes the study of multicriteria decision-making. In the Table below, some specific applications of supernetworks are given, for which results have been published in the literature. Subsequently, I elaborate upon several of the applications. For publications and additional references, see http://supernet.som.umass.edu

Examples of Supernetwork Applications
Telecommuting/Commuting Decision-Making
Teleshopping/Shopping Decision-Making
Supply Chain Networks with Electronic Commerce
Financial Networks with Electronic Transactions
Environmental and Energy Networks
Knowledge and Social Networks
Integrated Social and Supply Chain Networks
Electric Power Supply Chains and Transportation Networks

In particular, the supernetwork framework allows one to formalize the alternatives available to decision-makers, to model their individual behavior, typically, characterized by particular criteria which they wish to optimize, and to, ultimately, compute the flows on the supernetwork, which may consist of product shipments, the number of travelers between origins and destinations, the volumes of financial flows, energy flows, as well as the associated "prices." Hence, the concern is with human *decision-making* and how the supernetwork concept can be utilized to crystallize and inform in this dimension.

Supply Chain Networks and Electronic Commerce

The study of supply chain network problems through modeling, analysis, and computation is a challenging topic due to the complexity of the relationships among the various decision-makers, such as suppliers, manufacturers, distributors, and retailers as well as the practical importance of the topic for the efficient movement of products. The topic is multidisciplinary by nature since it involves particulars of manufacturing, transportation and logistics, retailing/marketing, as well as economics. In today's world, there is growing uncertainty and risk due to various threats and even illnesses which have affected dramatically the timely delivery of goods and have impacted transportation of humans as well.

The introduction of electronic commerce has, however, unveiled new opportunities in terms of research and practice in supply chain analysis and management since electronic commerce (e-commerce) has had a huge effect on the manner in which businesses order goods and have them transported with the major portion of e-commerce transactions being in the form of business-to-business (B2B). Estimates of B2B electronic commerce ranged from approximately 0.1 trillion dollars to 1 trillion dollars in 1998 and with forecasts reaching as high as $4.8 trillion dollars in 2003 in the United States. It has been emphasized that the principal effect of business-to-business (B2B) commerce, estimated to be 90% of all e-commerce by value and volume, is in the creation of new and more profitable supply chain networks.

In Figure SN-2 I depict a four-tiered supply chain network in which the top tier consists of suppliers of inputs into the production processes used by the manufacturing firms (the second tier), who, in turn, transform the inputs into products which are then shipped to the third tier of decision-makers, the retailers, from whom the consumers can then obtain the products. Here we allow not only for physical transactions to take place but also for virtual transactions, in the form of electronic transactions via the Internet to represent electronic commerce. In the supernetwork framework, both B2B and B2C can be considered, modeled, and analyzed. The decision-makers may compete independently across a given tier of nodes of the network and cooperate between tiers of nodes. In particular, my colleagues and I in a 2002 article in *Netnomics* applied the supernetwork framework to supply chain networks with electronic commerce in order to predict product flows between tiers of decision-makers as well as the prices associated with the different tiers. We assumed that the manufacturers as well as the retailers are engaged in profit-maximizing behavior whereas the consumers seek to minimize the costs associated with their purchases. The model therein determines the volumes of the products transacted electronically or physically.

As mentioned earlier, supernetworks may also be multilevel in structure. In [21], we demonstrated how supply chain networks can be depicted and studied as multilevel networks in order to identify not only the product shipments but also the financial flows as well as the informational ones. In Figure SN-3, I illustrate how a supply chain can be depicted as a multilevel supernetwork in which the financial network as well as the actual physical transportation network are also represented.

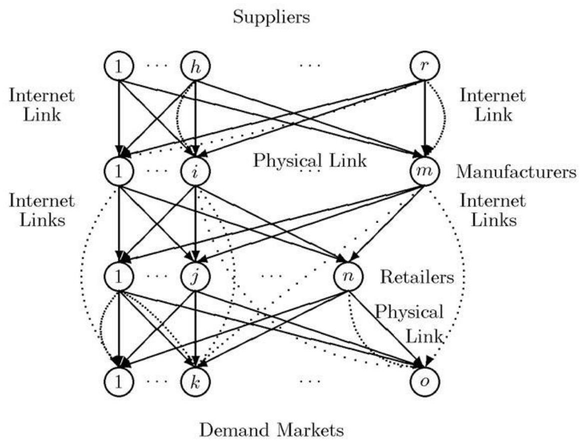


Figure SN-2. A Supply Chain Network (▷ Larger picture)

For example, in the supernetwork depicted in Figure SN-3, the logistical network affects the flows on the actual transportation network whereas the financial flows are due to payments as they proceed up the chain and as the transactions are completed. The information flows, in turn, are in the form of demand, cost, and flow data at the instance in time.

Obviously, in the setting of supply chain networks and, in particular, in global supply chains, there may be much risk and uncertainty associated with the underlying functions. Some research along these lines has already yielded promising results [19]. Continuing efforts to include uncertainty and risk into modeling and computational efforts in a variety of supernetworks and their applications is of paramount importance given the present economic and political climate.

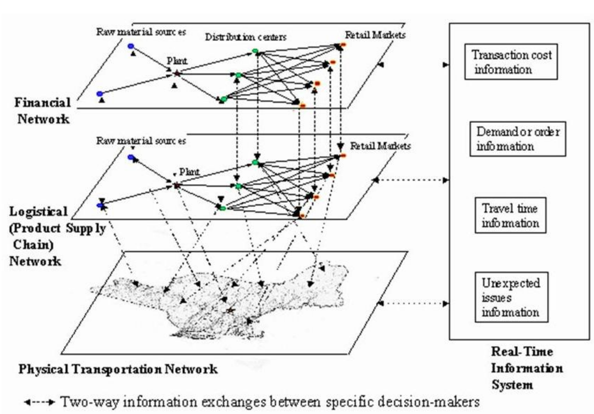


Figure SN-3. Supply Chain-Transportation Supernetwork Representation (▷ Larger picture)

In addition, I emphasize that the inclusion of environmental variables and criteria is also an important topic for research and practice in the context of supply chain networks, as has been demonstrated recently by Nagurney and Toyasaki [28] and is being presently investigated by my group in the context of electric power networks [38] and is generating much interest internationally.

Financial Networks and Electronic Transactions

Financial networks have been utilized in the study of financial systems since the work of Quesnay [31] in 1758, who depicted the circular flow of funds in an economy as a network. His conceptualization of the funds as a network, which was abstract, is the first identifiable instance of a supernetwork.

Quesnay's basic idea was subsequently applied in the construction of flow of funds accounts, which are a statistical description of the flows of money and credit in an economy. However, since the flow of funds accounts are in matrix form, and, hence, two-dimensional, they fail to capture the behavior on a micro level of the various financial agents/sectors in an economy, such as banks, households, insurance companies, etc. Moreover, the generality of the matrix tends to obscure certain structural aspects of the financial system that are of continuing interest in analysis, with the structural concepts of concern including those of financial intermediation.

Advances in telecommunications and, in particular, the adoption of the Internet by businesses, consumers, and financial institutions have had an enormous effect on financial services and the options available for financial transactions. Distribution channels have been transformed, new types of services and products introduced, and the role of financial intermediaries altered in the new supernetworked landscape. Furthermore, the impact of such advances has not been limited to individual nations but, rather, through new linkages, has crossed national boundaries.

The topic of *electronic* finance has been a growing area of study, as described in [18]. This is due to its increasing impact on financial markets and financial intermediation, and the related regulatory issues and governance. Of particular emphasis has been the conceptualization of the major issues involved and the role of networks is the transformations.

Now, I briefly describe a supernetwork framework for the study of financial decision-making in the presence of intermediation and electronic transactions. Further details can be found in [22,23]. The framework is sufficiently general to allow for the modeling, analysis, and computation of solutions to such problems.

The financial network model consists of: agents or decision-makers with sources of funds, financial intermediaries, as well as consumers associated with the demand markets. In the model, the sources of funds can transact directly electronically with the consumers through the Internet and can also conduct their financial transactions with the intermediaries either physically or electronically. The intermediaries, in turn, can transact with the consumers either physically in the standard manner or electronically. The depiction of the network at equilibrium is given in Figure SN-4.

It is assumed that the agents with sources of funds as well as the financial intermediaries seek to maximize their net revenue (in the presence of transaction costs) while, at the same time, minimizing the risk associated with the financial products. The solution of the model yields the financial flows between the tiers as well as the prices. We also allow for the option of having the source agents not invest a part (or all) of their financial holdings.

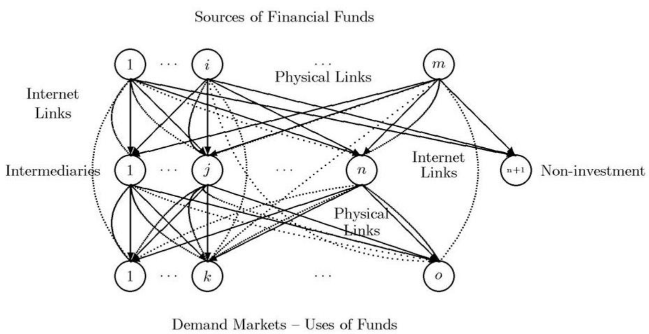


Figure SN-4. Financial Network (▷ Larger picture)

The Supernetwork Structure Reveals Answers to Questions Dating Back Half a Century

More recently, Liu and Nagurney [14] demonstrated that the supernetwork framework can also be used to show that financial equilibrium problems with intermediation can be reformulated and solved as transportation network equilibrium problems. Similarly, Wu et al. [38] proved that, as hypothesized in Chapter 5 of [1], electric power distribution networks can be reformulated and solved as transportation network equilibrium problems over an appropriately constructed abstract network or supernetwork. Copeland [5] asked whether "money flows like water or electricity?" In [14] we established that money flows like transportation flows and in [38] that electric power flows like transportation flows. Such reformulations of financial network problems and electric power supply chains have yielded, through the supernetwork concept, new interpretations of the governing equilibrium conditions in terms of path flows and associated costs. Because of such reformulations we now can apply additional computational procedures that yield path flow information (for a recent large-scale empirical application [15]).

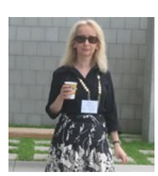
The Possibilities

In this essay, I have argued for the Supernetwork paradigm as a powerful tool for the study of network systems, emphasizing that it can capture not only the interrelationships among networks but, most importantly, the effects of human decision-making on the induced flows and prices. Through the computation of the flows and prices one can determine the optimal/equilibrium network designs and structures as well as the associated vulnerabilities. Hence, supernetworks provide not only powerful engineering and operations research/management science tools but also financial and economic ones. Finally, the supernetwork paradigm uniquely captures the human aspects and brings a richness to the conceptualization and the understanding of the underlying processes.

For example, the complex network literature [29] initiated principally by physicists, is only about a decade old, whereas the publications in OR/MS on networks date back a half century. An important aspect of the complex network physics literature concerns network efficiency measurement and vulnerability analysis. By showing that it is not just the topology that matters, but, also the associated costs and flows on links subject to the decision-makers' behavior, we have been able to generalize the results of Latora and Marchiori [12, 13]. In particular, we have been able to show the importance and the rankings of network components, i.e., the nodes and links in a more coherent and reasonable manner [26, 27].

In addition, the connections that have been established through the supernetwork paradigm of various applications are now being exploited to address the dynamics of such network systems, leading not only to new theories associated with equilibria, but with disequilibria, as well as time-dependent equilibria [14, 25]. For example, the well-known Braess paradox [2] (see also [3]), which has been the subject of much attention recently by the computer scientists [32] has been extended to a time-dependent version in which it was demonstrated that the paradox occurs only for a certain range of demands and, after a certain demand the new route is never used! Hence, if one considers network design issues in either telecommunication networks or in transportation networks, operating in a user-optimizing manner, then the addition of a new road/link may lead to increased user costs within a range of demands and the new route may not even be used past a certain demand; suggesting that the new link, which induces a new route, should not have been built since, over time, we may expect an increase in demand.

Supernetworks have, thus far, enabled the identification of similar structures and relationships between financial networks and transportation networks; electric power networks and transportation networks; supply chains and transportation networks, as well as transportation and telecommunication networks. Supernetworks have also been used to extend network-based results in other disciplines, including physics. Future research, I suspect, will include new concepts for dynamics, besides the recently formulated double-layered dynamics, new research into contagion, as well as into resiliency and robustness of network systems.



Anna takes a break at the Puerto Rico meeting.

References

[1] D. Boyce, H.S. Mahmassani, and A. Nagurney, A Retrospective on Beckmann, McGuire, and Winsten's *Studies in the Economics of Transportation*, *Papers in Regional Science* 84:1 (2005), 85–103.

[2] D. Braess, Uber ein Paradoxon der Verkehrsplanung, *Unternehmenforschung* 12 (1968), 258–268.

[3] D. Braess, A. Nagurney, and T. Wakolbinger, On a Paradox of Traffic Planning, Translation of the (1968) Original D. Braess Paper from German to English, *Transportation Science* 39:4 (2005), 446–450.

[4] M.-J. Cojocaru, P. Daniele, and A. Nagurney, Double-Layered Dynamics: A Unified Theory of Projected Dynamical Systems and Evolutionary Variational Inequalities, *European Journal of Operational Research* 175:1 (2006), 494–507.

[5] M.A. Copeland, *A Study of Moneyflows in the United States*, National Bureau of Economic Research, New York, 1952.

[6] S. Dafermos, The Traffic Assignment Problem for Multimodal Networks, *Transportation Science* 6:1 (1972), 73–87.

[7] S. Dafermos, Integrated Equilibrium Flow Models for Transportation Planning, in M.A. Florian (Ed.), *Traffic Equilibrium Methods, Lecture Notes in Economics and Mathematical Systems* 118, Springer-Verlag, NY, 1976, 106–118.

[8] S. Dafermos and F.T. Sparrow, The Traffic Assignment Problem for a General Network, *Journal of Research of the National Bureau of Standards* 73B (1969), 91–118.

[9] P.J. Denning, The Science of Computing: Supernetworks, *American Scientist* 73 (1985), 127–129.

[10] J. Fallows, The Java Theory, *The Atlantic Monthly* 277:3 (1996), 113–117.

[11] Illinois State Bar Association, Opinion No. 96–10, Topic: Electronic Communications; Confidentiality of Client Information; Advertising and Solicitation, May 16, 1997.

[12] V. Latora and M. Marchiori, Efficient Behavior of Small-World Networks, *Physical Review Letters* 87:19 (2001), Article No. 198701 (4 pp).

[13] V. Latora and M. Marchiori, Economic Small-World Behavior in Weighted Networks, *European Physical Journal B* 32:2 (2003), 249–263.

[14] Z. Liu and A. Nagurney, Financial Networks with Intermediation and Transportation Network Equilibria: A Supernetwork Equivalence and Reinterpretation of the Equilibrium Conditions with Computations, *Computational Management Science* 4:3 (2007), 243–281.

[15] Z. Liu and A. Nagurney, *An Integrated Electric Power Supply Chain and Fuel Market Network Framework: Theoretical Modeling with Empirical Analysis for New England*, Isenberg School of Management, University of Massachusetts, Amherst, MA, 2007.

[16] A. Nagurney, *Network Economics: A Variational Inequality Approach*, 2nd edition, Kluwer Academic Publishers, The Netherlands, 1999.

[17] A. Nagurney, Navigating the Network Economy, *OR/MS Today* 27:3 (2000) 74–75.

[18] A. Nagurney, Editor, *Innovations in Financial and Economic Networks*, Edward Elgar Publishing, UK, 2003.

[19] A. Nagurney, J. Cruz, J. Dong, and D. Zhang, Supply Chain Networks, Electronic Commerce, and Supply Side and Demand Side Risk, *European Journal of Operational Research* 164:1 (2005), 120–142.

[20] A. Nagurney and J. Dong, *Supernetworks: Decision-Making for the Information Age*, Edward Elgar Publishing, UK, 2002.

[21] A. Nagurney, K. Ke, J. Cruz, K. Hancock, and F. Southworth, Dynamics of Supply Chains: A Multilevel (Logistical/Informational/Financial) Network Perspective, *Environment and Planning B* 29:6 (2002), 795–818.

[22] A. Nagurney and K. Ke, Financial Networks with Intermediation, *Quantitative Finance* 1:4 (2001), 441–451.

[23] A. Nagurney and K. Ke, Financial Networks with Electronic Transactions: Modeling, Analysis, and Computations, *Quantitative Finance* 3:2 (2003), 71–87.

[24] A. Nagurney, J. Loo, J. Dong, and D. Zhang, D., Supply Chain Networks and Electronic Commerce: A Theoretical Perspective, *Netnomics* 4:2 (2002), 187–220.

[25] A. Nagurney, D. Parkes, and P. Daniele, The Internet, Evolutionary Variational Inequalities, and the Time-Dependent Braess Paradox, *Computational Management Science* 4:4 (2007), 355–375.

[26] A. Nagurney and Q. Qiang, A Network Efficiency Measure for Congested Networks, *Europhysics Letters* 79:3 (2007), Article 38005 (5 pp).

[27] A. Nagurney and Q. Qiang, A Network Efficiency Measure with Application to Critical Infrastructure Networks, *Journal of Global Optimization*, in press.

[28] A. Nagurney and F. Toyasaki, Reverse Supply Chain Management and Electronic Waste Recycling: A Multitiered Network Equilibrium Framework for E-Cycling, *Transportation Research E*, 41:1 (2005), 1–28.

[29] M. Newman, The Structure and Function of Complex Networks, *SIAM Review* 45:2 (2003), 167–256.

[30] A. Noveen, V. Hartenstein, C.-M. Chuong, Gene Networks and Supernetworks: Evolutionarily Conserved Gene Interactions, in C.-M. Chuong (Ed.), *Molecular Basis of Epithelial Appendage Morphogenesis*, Landes Bioscience, Austin, TX, 1998, pp. 371–391.

[31] F. Quesnay, *Tableau Économique*, 1758. Reproduced in Facsimile with an Introduction by H. Higgs by the British Economic Society, 1895.

[32] T. Roughgarden, *Selfish Routing and the Price of Anarchy*, MIT Press, Cambridge, MA, 2005.

[33] L. Schubert, R. Goebel, and N. Cercone, The Structure and Organization of a Semantic Net for Comprehension and Inference, in N.V. Findler (Ed.), *Associative Networks: Representation and Use of Knowledge by Computers*, Academic Press, New York, NY, 1979.

[34] Y. Sheffi, *Transportation Network Equilibrium with Discrete Choice Models*, Ph. D. Thesis, Civil Engineering Department, Massachusetts Institute of Technology, Cambridge, MA, 1978.

[35] Y. Sheffi, *Urban Transportation Networks —- Equilibrium Analysis with Mathematical Programming Methods*, Prentice-Hall, Englewood Cliffs, NJ, 1985.

[36] Telecom 95 (1995), MCI and the Internet: Dr. Vinton G. Cerf, Keynote Address, http://www.itu.int/TELECOM/wt95/pressdocs/papers/cerf.html.

[37] J.G. Wardrop, Some Theoretical Aspects of Road Traffic Research, *Proceedings of the Institute of Civil Engineers, Part II*, 1952, 325–378.

[38] K. Wu, A. Nagurney, Z. Liu, and J.K. Stranlund, Modeling Generator Power Plant Portfolios and Pollution Taxes in Electric Power Supply Chain Networks: A Transportation Network Equilibrium Transformation, *Transportation Research Part D*, 11:3 (2006), 171–190.

Data Mining in Industrial Applications and Innovation

Andrew Kusiak ▷
andrew-kusiak@uiowa.edu

Andrew Kusiak is Professor of Mechanical and Industrial Engineering at The University of Iowa, and the Research Coordinator of their Intelligent Systems Laboratory. He has championed *Innovation Science* in the manufacturing environment, having just published a primer in *International Journal of Computer Applications in Technology* 28:2-3, 2007, 140–149. Dr. Kusiak serves on journal editorial boards, edits book series, and is the Editor-in-Chief of the *Journal of Intelligent Manufacturing*. He is also the Past Chair of the INFORMS Data Mining Section.

About Data Mining

Data mining [23] has more than a decade-old history, yet it remains a mystery to many. There is a range of intuitive opinions of data mining, ranging from considering it a branch of statistics to a somewhat deterministic tool. These views are in essence correct as data mining has different shades by encompassing algorithms that are deeply versed in statistics, computational intelligence, and those resembling deterministic approaches.

How Different is Data Mining from Traditional OR/MS Methods?

If statistics is your basic tool of interest, it is worthy remembering that statistical models describe populations of objects. Data mining models could describe a population as well; however, the population-based field might be too crowded for data mining to enter. Data mining has a better chance to distinguish itself as an individual-based approach rather than a population-based science.

Statistical models, e.g., regression, use all parameters specified by the user. It is true that not all of them are equally significant and some may not be incorporated in the model. However, while using the model all independent parameter values need to be provided to determine the decision (the value of the dependent variable). A decision tree or decision rules extracted with a decision-tree or a decision-rule algorithm do not require all values of independent variables for making decisions. Data mining algorithms usually have the ability to select important parameters; however, an entire branch of data mining known as feature selection deals in refined ways of feature (parameter) selection.

While statistical models largely evolve around independent and statistically significant variables, data mining algorithms can make good use of insignificant and even correlated variables. Combinations of insignificant variables often turn into powerful decision-making models.

For those who enjoy using operations research, data mining brings a fresh perspective. An operations research analyst carefully studies the problem on hand and depending on her/his background recommends a particular model and an algorithm. The model could be deterministic or stochastic. By recommending (fitting) a model or a collection of models the OR analyst fixes the number of variables, e.g., a transportation model uses specific variables, constraints, and the objective function. It is true that the selected model can be modified by adding new variables, constraints, and so on. Often the model is built from scratch; however, when fully developed, usually the structure resembling the traditional OR models can be identified. As statisticians enjoy curve fitting, operations researchers fit models or their elements and we call it modeling. In some cases, e.g., the formal modeling phase is replaced with algorithm-fitting or its development.

Another issue deserving attention is that operations research deals almost exclusively with quantitative variables. The discussion on combining quantitative and qualitative variables in operations research modeling has been lasting decades without much success. Data mining might have solved the quantitative and qualitative variable dilemma of operations research. Many data mining algorithms handle with great ease strings of numbers and qualitative values. In fact, an entire sentence could be one of many variables considered in data mining. Data mining allows casting a wide net over all possible quantitative and qualitative variables. In addition, unlike operations research, where modeling is in essence top-down (model fitting), in data mining a bottom-up approach is followed. A typical data mining algorithm derives a model from the data rather than trying to fit a preconceived model. One could say that a traditional modeler comes with luggage (e.g., dependent on previous training and experience), while a data miner does not. A machine learning algorithm extracts the model.

Where Data Mining Works Best?

A data mining algorithm does only a partial job, i.e., it creates a model. In cases when a user is looking for patterns the results produced by the data mining algorithm may suffice. This type of data mining is called *descriptive*. The latter does not provide a competitive edge for data mining in areas where statistics and operations research are dominant, i.e., decision making. A data-mining algorithm derives a model (or modes) for a purpose. If the purpose of data mining is to make decisions, then decision-making algorithms are needed. Decision making with the data-mining derived models is most often rushed over, if not ignored. The machine learning community may think of decision making in terms of statistical tests, e.g., cross-validation or boot strapping. While such tests are useful in comparing performance of various machine learning algorithms, they could not serve industrial and service applications.

When one moves from the realm of data mining limited to a machine learning task to actual applications, the specific algorithm selected for knowledge extraction may even become secondary to the decision-making algorithm. There is no doubt that the area of decision making in the context of data mining needs attention. Otherwise, the discussion on the utility of data mining will continue.

Illustrative Applications of Data Mining

Many books have been written on applications of data mining and all cite plenty of examples. It is safe to say that whenever there is historical data, its analysis with data mining could be beneficial. Rather than presenting a long list of potential data mining applications, illustrative applications developed by one laboratory, the Intelligent Systems Laboratory at the University of Iowa, will be presented. The presented examples make only a small subset of data mining projects conducted in almost a decade. These projects could be classified into three categories, as follows.

1. Medical applications

   Genetics [17,18]. Projects in genetics are often sponsored by pharmaceutical companies. Do not expect publications based on company-collected generic data as it is usually closely protected.

   Medical diagnosis (e.g., lung cancer [11], bladder cancer [19]). The data sets generally come from CROs (clinical research organizations) and physicians involved in large-scale medical projects.

   Dialysis [10]. This area offers plenty of unanswered clinical questions. The best data sets came from a software company formed outside hospitals and collecting data at many different clinics.

   Heart disease [9,13]. The project goals were to predict health status after heart surgery and medical interventions of an infant suffering from a genetic disorder known as the *hypoplastic left heart syndrome*. It has taken more than two years to develop a data collection systems hooked up to the intensive care unit equipment. The screen shot of the developed system is shown in Figure DM-1.

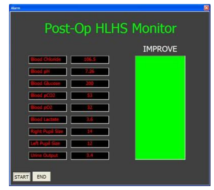


Figure DM-1. Infant Health Status-Predicting System (▷ Larger picture)

2. Manufacturing applications

   Semiconductor industry (quality improvement applications [4,7]). The data was generated by a semiconductor company.

   Electronics assembly (product fault detection [12,14]). The data was generated by a company.

   Mass customization [3,14]. A comprehensive tool for configuring products was developed for over a hundred product families. A large number of data sets were web and ERP (enterprise resource planning) system generated and provided by a company. One of many screens of the system implemented in industry is shown in Figure DM-2.

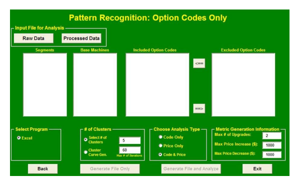


Figure DM-2. The Mass Customization System (▷ Larger picture)

3. Energy applications

   Water chemistry [15]. The data was provided by a private electric utility company. It was impressive to watch a data set of 60,000 observations compressing to a few decision rules.

   Combustion efficiency [8,16,20]. Electric utility companies collect large volumes of data. A system for maximizing combustion efficiency was developed (Figure DM-3).

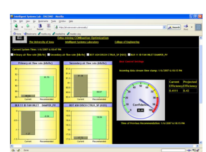


Figure DM-3. The System DACOMO for Optimizing Energy Efficiency (▷ Larger picture)

   Equipment diagnostics. The research on combustion efficiency has led to the development of a system for equipment diagnostics. Some faults are sensed by the system hours or longer before actually happening.

Future Prospects of Data Mining?

It is going to take years before even areas indicated by the above sample projects will be saturated with data mining. Rather than considering many directions that data mining may take, one area — innovation — will be considered. In fact, data mining may become a significant component of a new science, the *science of innovation* [5,6].

Innovation enables organizations to compete effectively [1]. The need to understand innovation has resulted in renewed interest among research and corporate communities.

Though numerous innovation studies have been published, the literature on innovation is filled with myths and inconclusive research findings. Innovation is often discussed based on experiences specific to a particular case study. For example, innovation undertakings at companies such as 3M and Apple have been broadly studied. However, it is not known to what degree these findings would produce similar results in other corporations.

Some companies have followed conventional thinking for too long [2]. Most widely cited examples include Microsoft's reluctance to embrace the concept of open-source software; Polaroid's slow progress on digital cameras; GM's and Ford's lack of enthusiasm for hybrid cars, improvements in fuel economy, and failure to embrace the common platform concept and media companies overlooking blogs.

What is Innovation?

Innovation is an iterative process aimed at the creation of new products, processes, knowledge or services by use of new or existing scientific knowledge. Some use the terms "technology-based innovation" or "technological innovation" to emphasize the role of technology.

Innovation Drivers

The traditional view of the innovation process is based on the *technology-push approach* — a linear model emphasizing the advancements in science and technology as a sole event triggering the creation of a new design. The process is initiated by a technological breakthrough and followed by a series of developments. This view is criticized as neglecting the influence of customers.

Companies use various means to reach out to customers to incorporate their needs into the product development process. Many researchers have suggested that companies use an incorrect approach and measurements when consulting with customers. Ulwick [21] pointed out that companies should not expect solutions to be offered by potential customers; rather, they should ask them about the desired product's characteristics. He argued that customers may only know what they have experienced and may have a limited frame of reference when suggesting innovative ideas. In addition, companies that link their products too closely to their customers may end up creating incremental innovation. Veryzer [22] emphasized the need for caution with customer input, and pointed out the importance of discontinuous product development, e.g., the customer's input should be introduced later in the project. Christensen [1] stated that customers may emphasize the product's functionality to too great a degree. For example, many customers buy milkshakes based on the drink's thickness and strong flavor. The milkshake is thus competing with the "functionality" of such complementary products as sandwiches, soft drinks, and salads. Without understanding this phenomenon, fast food companies may develop a product that is completely at odds with what the customers actually want.

The Need for Requirements-driven Innovation

Recognizing customer requirements in development of products and services has been widely practiced by successful corporations. Traditionally, requirements have been managed by marketing departments. Conventional marketing techniques such as questionnaires, focus groups or interviews were widely used. In the digital world, customer opinions are recorded on blogs, social network forums, and other digital media. Using digital media to generate useful information about customer requirements is of paramount importance. Google's business model is based on matching advertisements with targeted audiences. More recent social network internet sites, such as MySpace or YouTube, follow Google's model by sending the right information to the right people.

Sources of Requirements

In the past two decades, the design of products and services has been largely driven by customers. After all, the customer buys a product or uses a service. The "customer-as-the-king" model was preceded by the "engineer-as-the-king" (often designer) model, in which technical experts made the decisions for the customer. The customer was expected to accept the offered product or service.

Both models of eliciting requirements have focused on the product and service functions. Product innovation calls for additional requirements, making it worthy of the label "innovative product." The sources of innovation-fostering requirements are much wider, and they include:

Customers. The information from the customers should be collected over the product's life-cycle rather than during a limited time frame. Processing that information and blending it with other sources of data and information could be the ultimate key to the success of the designed product.

Domain experts. Though the importance of the voice of the engineer in forming requirements has been marginalized in the last few decades, it needs to be brought back and expanded when innovating. It is true that a customer is the one who ultimately pays for the product; however, he may not be aware of the possibilities that a new technology or a product/process combination may offer. A technologist may generate innovative features of a product.

Legacy materials. All kinds of standard and digital libraries could be searched in the quest of innovation. The search would involve hypotheses, theories, innovation rules, and information about inventors and innovators. Data-mining algorithms could create previously unseen value in fusing data and information from various sources.

Product life-cycle data. A product leaves a data trail over its life cycle. This is in addition to the information provided by the customers or experts before and after the product has entered the market. The volume of data collected can be large, e.g., imagine a database of cockpit and maintenance data collected over the useful life of an airplane. The product's lifetime data can deliver valuable knowledge leading to requirements spurring innovation.

Having outlined the role of requirements in innovation as a "data generator," the role of data mining in this exciting undertaking is obvious. It is going to be used to discover patterns leading to market acceptance of candidate solutions.

Without going into details, many agree that market relevance and market acceptance distinguish innovation from invention and creation (see Figure DM-4).

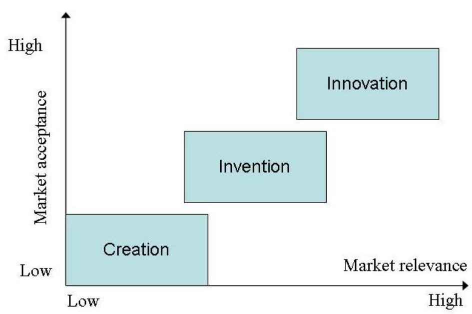


Figure DM-4. Relationships Among Creation, Invention, and Innovation

The market determines whether a creation or an innovation becomes an innovation. The market acceptance and relevance can be expressed in economic terms (e.g., market share, profit) or using other metrics (e.g., social acceptance). Data mining is likely to play a key role in focusing on and pursuing creations and inventions that have a high likelihood of becoming innovations.

Summary

This article presented a view of data mining developed and largely implemented across numerous domains. Data mining offers plenty of space to different views and levels of intellectual and experiential endeavors.

The journey of data mining has only begun and will continue for decades to come. Changes in names and directions may happen on the way, however, the spirit of knowledge extraction to serve a purpose will certainly prevail. For the data mining community within INFORMS that purpose is destined to be decision making. The rate of growth of activities around the INFORMS Data Mining Section over the recent years is one of the many sources of the optimism. The data mining activities within INFORMS have been enjoying a strong corporate support, which may make the section unique.

Andrew takes his wife, Anna, to a data mining meeting. It's a tough job, but someone has to do it.

References

[1] C.M. Christensen, *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*, Harvard Business School Press, Boston, MA, 1997.

[2] D.A. Garvin and L.C. Levesque, Meeting the challenge of corporate entrepreneurship, *Harvard Business Review* 84:10 (2006), 102–112.

[3] A. Kusiak, M.R. Smith, and Z. Song, Planning Product Configurations Based on Sales Data, *IEEE Transactions on Systems, Man, and Cybernetics:* Part C 37:4 (2007), 602–609.

[4] A. Kusiak, Decomposition in Data Mining: An Industrial Case Study, *IEEE Transactions on Electronics Packaging Manufacturing* 23:4 (2000), 345–353.

[5] A. Kusiak, Innovation: The Living Laboratory Perspective, *Computer-Aided Design and Applications* 4:6 (2007), 863–876.

[6] A. Kusiak, Innovation Science: A Primer, *International Journal of Computer Applications in Technology* 28:2–3 (2007), 140–149.

[7] A. Kusiak, Rough Set Theory: A Data Mining Tool for Semiconductor Manufacturing, *IEEE Transactions on Electronics Packaging Manufacturing* 24:1 (2001), 44–50.

[8] A. Kusiak, A. Burns, S. Shah, and N. Novotny, Detection of Events Causing Pluggage of a Coal-Fired Boiler: A Data Mining Approach, *Combustion Science and Technology* 177:12 (2005), 2327–2348.

[9] A. Kusiak, C.A. Caldarone, M.D. Kelleher, F.S. Lamb, T.J. Persoon, and A. Burns, Hypoplastic Left Heart Syndrome: Knowledge Discovery with a Data Mining Approach, *Computers in Biology and Medicine* 36:1 (2006), 21–40.

[10] A. Kusiak, B. Dixon, and S. Shah, Predicting Survival Time for Kidney Dialysis Patients: A Data Mining Approach, *Computers in Biology and Medicine* 35:4 (2005), 311–327.

[11] A. Kusiak, J.A. Kern, K.H. Kernstine, and T.L. Tseng, Autonomous Decision-Making: A Data Mining Approach, *IEEE Transactions on Information Technology in Biomedicine* 4:4 (2000), 274–284.

[12] A. Kusiak and C. Kurasek, Data Mining of Printed-Circuit Board Defects, *IEEE Transactions on Robotics and Automation* 17:2 (2001), 191–196.

[13] A. Kusiak, I.H. Law, M.D. Dick, The G-Algorithm for Extraction of Robust Decision Rules: Children's Postoperative Intra-

atrial Arrhythmia Case Study, *IEEE Transactions on Information Technology in Biomedicine* 5:3 (2001), 225–235.

[14] A. Kusiak and M. Smith, Data Mining in Design of Products and Production Systems, *IFAC Annual Reviews in Control* 31:1 (2007), 147–156.

[15] A. Kusiak and S. Shah, A Data-Mining-Based System for Prediction of Water Chemistry Faults, *IEEE Transactions on Industrial Electronics* 53:2 (2006), 593–603.

[16] A. Kusiak and Z. Song, Combustion Efficiency Optimization and Virtual Testing: A Data-Mining Approach, *IEEE Transactions on Industrial Informatics* 2:3 (2006), 176–184.

[17] S. Shah and A. Kusiak, Cancer Gene Search with Data-Mining and Genetic Algorithms, *Computers in Biology and Medicine* 37:2 (2007), 251–261.

[18] S.C. Shah and A. Kusiak, Data Mining and Genetic Programming Based Gene/SNP Selection, *Artificial Intelligence in Medicine* 31:3 (2004), 183–196.

[19] S. Shah, A. Kusiak, and M. O'Donnell, Patient-Recognition Data-Mining Model for BCG-plus Interferon Immunotherapy Bladder Cancer Treatment, *Computers in Biology and Medicine* 36:6 (2006), 634–655.

[20] Z. Song and A. Kusiak, Constraint-Based Control of Boiler Efficiency: A Data-Mining Approach, *IEEE Transactions on Industrial Informatics* 3:1 (2007), 73–83.

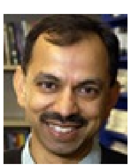
[21] A.W. Ulwick, Turn Customer Input into Innovation, *Harvard Business Review* 80:1 (2002), 91–97.

[22] R. Veryzer, The roles of marketing and industrial design in discontinuous new product development, *Journal of Product Innovation Management* 22:1 (2005), 22–41.

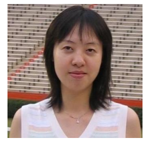
[23] I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd Edition, Morgan Kaufmann, 2005.

Search-Engine Advertising: Dynamic Auctions under Performance-Based Pricing

Hemant K. Bhargava ▷, hemantb@ucdavis.edu
Juan (Jane) Feng ▷, jane.feng@cba.ufl.edu



Hemant K. Bhargava is Professor of Management and of Computer Science at The University of California at Davis. He received his Ph. D. and M.S. from The Wharton School in Decision Sciences in 1990. His recent work on quality uncertainty in the context of IT services and products explains how firms can employ contingency pricing schemes in order to better manage the effects of quality uncertainty and improve their competitive position. His research interest in the economics of information systems and IT industry covers sponsored search in recommenders and search engines, stockout compensation practices, performance-based pricing for IT goods, product versioning, pricing strategies for grid computing, and information structures in contract design. Dr. Bhargava received the Menneken Prize for Excellence in Research at the Naval Postgraduate School, and his research on *DecisionNet* won best-paper awards at leading information systems conferences.



Juan Feng is Assistant Professor of Information Systems & Operations Management at the University of Florida, Warrington College of Business. In 2003, she received her doctorate degree in Management Science & Information Systems from Penn State University, with a dual title in Operations Research. In 2002, she was the Penn State representative to the *Doctoral Student Consortium*, held in Barcelona. That same year Jane received the e-Business Research Center's *Doctoral Support Award*. Jane's current research is in economics of information systems, sponsored search auctions in the information gatekeepers, and pricing and competition strategies.

In the last 15 years, the Internet has spurred on a vast collection of online information services — real-time traffic, weather, news, stocks, audio and video streaming, travel search, shopbots, recommendation services, photo and video sharing, social networking. And the biggest gorilla of them all, Internet search engines (the leading search engine, Google, had an average of 140,000,000 searches per day during August 2007[†]). The dominant revenue source for search engines and many other online information services is advertising by merchants who hope to reap monetary rewards through trade with users of these services. This form of advertising is commonly referred to as "sponsored search," capturing the idea that some of the search results (or, generally, information content) returned by these services are paid by third-party sponsors: communication of these "relevant" or targeted messages enables the search engine to offer a free service to end-users. Other popular industry terms for this practice are paid search and preferential placement. The first use of sponsored results (in 1998) is credited to the search engine GoTo.com, which later becomes Overture Services, which was acquired by Yahoo! in 2003. Sponsored search is widely accepted to be the driving factor behind the meteoric rise in commercial value of the current industry leader, Google. It is behind the multi-billion dollar valuations of popular social networking sites, and has also entered the realm of inter-personal phone calls: recently, both wireless and fixed-line telephone operators have begun offering free service in exchange for advertising messages determined by monitoring the conversation. While sponsored results are observed in many information services, this article uses the context of *search engines* for ease of exposition.

Sponsored search advertising, having hatched over a foundation of modern information technologies, has several innovative characteristics that make it very data, computation, and algorithm intensive. First, the advertisements (or "messages") are contextual rather than general and non-personal. The mapping from search query to advertisement is made in real-time based on contextual information from the current search as well as past behavior. Second, advertisements tend to be very fine-grained (e.g., for a specific product out of tens of thousands that a merchant may carry, vs. simply advertising the brand). Merchants therefore have to manage their advertising expenditures over thousands of messages. The search engines, in turn, manage inventories of hundreds of thousands

[†] Nielsen NetRatings, 2007, as quoted on SearchEngineWatch.com.

of keywords or search phrases. Third, the advertising price for each keyword is discovered mostly through online, repetitive, auctions. The search engine announces a mechanism or format for the auction — a separate auction is conducted for each keyword — and merchants must place bids on each keyword they wish to compete on (these bids are considered independent, even though merchants' preferences may be interrelated). Fourth, the advertising payment is "performance-based" meaning that the actual fee is based not just on impressions of the ad (as in traditional advertising) but on some measure of advertising performance (the commonly used measure is "clickthrough" but firms are experimenting with other measures of conversion such as page-views and product sales). Fifth, for each keyword, the merchants' bids and the search engines' selection of winners — and more generally, the choice of auction format — must be determined under uncertainty about one crucial set of variables, the performance of each merchant's ad.

Collectively, these five characteristics lead to several novel and complex optimization problems. Let us begin with a simplified statement of this practice. Consider a single auction for a single sponsored slot on a single keyword. I merchants are interested in contacting consumers who search on this keyword. The search engine awards the slot in order to maximize its expected value. Under performance-based pricing, the expected revenue from awarding the slot to merchant i is a function of merchant i's clickthrough rate, which the search engine can (imperfectly) estimate, and the merchant's bids. This appears like a straightforward expected value maximization problem: compute all the expected revenues and pick the highest one. However, in reality, the same keyword is searched thousands of times a day or week. If the search engine's initial estimates of clickthrough rates are incorrect (as is likely the case), a myopic allocation scheme (award all exposures to the merchant with highest expected revenue) is predictably suboptimal. This raises the need to refine the estimates of clickthrough rates. But such learning requires the search engine to provide some exposures to merchants that have, for that particular exposure, low expected revenue. This creates a tradeoff between *exploring* and *exploiting* available information. How to best allocate the slots over a number of searches in order to optimally balance exploration and exploitation? Readers will readily recognize the multi-arm bandit problem here. The auction design and analysis is further complicated because each search is an opportunity to allocate not just one but several slots. These slots are not identical, rather they are vertically differentiated variants (its safe to assume that all merchants agree that the top-ranked slot garners more user attention than lower-ranked slots). Lower-ranked slots are subject to the "law of declining clickthrough rates". This adds to the complexity of revenue maximization in the single auction problem. Moreover, this entire analysis depends on how merchants respond to the auction rules set by the search engine. Merchants' bidding strategy, the allocation of the sponsored slots, and the search engine's revenues can differ substantially based on the auction format, the analysis of which is in turn complicated due to the factors mentioned above. The rest of this article elaborates on the problem characteristics and challenges outlined above.

Auction Design: How and Whom to Allocate Sponsored Slots?

Online auctions have become the dominant method for allocating sponsored slots. Consider again a single auction for a single sponsored slot on a single keyword, of value to I merchants. Suppose merchant i has value b_i for each customer contact directed by the search engine; this is the merchant's reservation price for the slot. This may be the average profit on a sale divided by the number of contacts needed to make a sale, or may represent the cost of customer acquisition using an alternative channel. Let B_i be the amount that merchant i *bids* in the auction, which depends on b_i and the auction format. The auction format determines how merchants bid, but we ignore the nuances of the winner determination rule and merchants' bidding strategy in order to focus on other complexities unique to this problem. Let p_i denote the search engine's estimate of merchant i's clickthrough probability (in practice, clickthrough rates typically range from 0 to 3%). Let R_i denote the price-per-click that the search engine can charge if it allocates the slot to merchant i. Then the expected payoff from awarding the slot to merchant i is p_iR_i. The search engine's auction design problem — how to select the slot winner, what to charge — poses several design choices.

The unit for charging. While advertising fees were traditionally linked to the number of impressions, internet-based advertising is dominated by performance-based pricing. The commonly used "performance" measure is simply a "click", but the industry is experimenting with more aggressive actions such as page views, downloads, and product sale [8]. For example, *ZiXXo* employs "pay-per-print" while *Ingenio* uses a "pay-per-call" model. The more "aggressive" the performance measure, the less likely that the search engine can accurately estimate or control the probability of successful performance. Hence the choice of unit for pricing has fundamental implications on other aspects of auction design.

Determining the winner. The simplest mechanism for ranking bidders is "rank by bid" (used in recent years by GoTo, Overture, and Yahoo!), which may be reasonable when the winner pays per impression. But under performance-based pricing, the slot generates revenue only when some user takes an action (such as a click), raising the need to weed out frivolous bids that may have high value but generate no click-through revenue (Yahoo! employed manual editorial filtering when they used this scheme). Recognizing this, some search engines sort the bidders by a product $V_i = p_iB_i$ which models the expected payoff from awarding the slot to merchant i (this approach has been employed by Google). V_i can be thought of as the value of merchant i to the search engine. Several papers have found that it is profitable to rank the winners by $p_i B_i$ [5, 10, 6].

Pricing the slot. In terms of the payment scheme (what price R to charge the winner), the "pay your bid" scheme (first-price auction, used in recent years by GoTo, Overture, and Yahoo!) requires the winner ($j = \arg\max_i V_i$) to pay B_j per click. Or, the search engine may use "pay an increment over next-highest bid". This would be straightforward in a "pay-per-impression" model (set R_j to the value $B_k + \epsilon$ where k is the next highest bidder after j). However, performance-based pricing introduces a small twist: should the rule be applied to bid amount B_i (the well-understood measure) or to expected revenue $p_i R_i$? These two might produce very different results because the p_i's may be substantially different across the advertisers, moreover the estimates may be incorrect. For example (focusing on *ex ante* analysis), suppose $p_1 = 0.04$ with $B_1 = 4$ (expected revenue \$0.16 per impression), while $p_2 = 0.01$ with $B_2 = 9$ (expected revenue \$0.09 per impression). Merchant 1 wins the slot, but how much should he pay per click? Applying "next highest bid plus a penny" to B_i's would yield \$10.01 (obviously meaningless because merchant 1 is willing to pay no more than \$4 per click), whereas applying the rule to expected revenue would require $R = \$2.50$ per click. But this second, more reasonable, approach also poses problems when there are multiple slots to allocate. Suppose $p_3 = 0.01$ with $B_3 = 7$ (expected revenue \$0.07 per impression). So, now, the second slot (which is awarded to merchant 2) would require the winner to generate an expected revenue of \$0.08, with a per-click price of \$8. Here, we have a seemingly strange outcome that the merchant in the top slot pays \$2.50 per click while one in the second slot pays a higher per-click fee of \$8.

A third alternative is to set R_j to be the minimum of the price obtained under the two approaches. This reduces the search engine's revenue but has desirable properties. It is simple and fair: the merchant that occupies rank i pays (a small increment) more per click than the merchant at rank $i+1$. The higher rank can produce substantially higher expected revenue because the slot (if awarded to merchants) is likely to generate more clicks. The property that, on a per click basis, a lower ranked merchant pays less than a higher-ranked one, reduces merchants' incentives for bid-shading.

The overall challenge in auction design is to develop a complete specification of the auction format such that it achieves some objective (maximize the search engine's profit, or the advertiser's payoff, or the consumer's welfare, or the total social welfare, etc.). However, deriving this complete specification and its associated equilibrium outcomes are non-trivial because (1) the number of slots available and the number of bidders participating in the auction are usually more than just a few, (2) advertisers from different industry and backgrounds may have quite different strategic objectives when participating in such auctions (for example, some advertisers may be interested in raising awareness/impression, and some may prefer to signal their high quality levels; some advertisers may prefer the top-rank position, while some may prefer may prefer a lower rank to attract more serious consumers, etc.), and (3) as a result, bidders may have quite different daily budgets to participate in these auctions. Moreover, there are dynamic features that are hard to completely capture in the commonly used static models: for example, new advertisers may participate in these auctions any time (while others advertisers have sufficiently long history and experience), and bidders have the flexibility of adjusting individual bids in real time. Thus, in reality, focusing purely on designing the optimal auction may not be sufficiently productive. Equally worthwhile is comparative analysis of intuitive designs or those observed in practice. Factors that deserve attention while studying an auction mechanism include the following.

- **Clickthrough rates and revenue at each position.** Experience in several domains (travel search, movies, listing guides) indicates that higher ranks get, *ceteris paribus*, more attention and click rates. Some studies of search auction design assume that the clicks received by an advertisement do not depend on the identity of the advertiser. However, for the user with a search problem, all merchant listings are *not* equal. The "more relevant" merchants are likely to get higher clicks at the same rank than less relevant ones. This raises the complexity of estimating expected revenues, winner determination, and pricing.
- **The advertiser's optimal bidding behavior with a certain budget, given he/she participates in a cluster of keywords.** It may be wise to adjust bids/budget based on the cost/ben-efit of each auction, which may include different competition status (current price for different ranks) of these keywords, and the click-throughs/conversions attracted by these keywords. This is a complicated problem because bidders may compete for different advertising positions for different keywords. In addition, in the presence of budget constraints, merchants might game their bids in a second-price auction: a lower-value bidder may have incentive to bid high in initial rounds (note that this does not actually cause it pay a higher price for its own slot) to quickly force its competition to run out of its daily budget and drop out, thereby enabling it to win the high slot for a low price in the future.
- **The outcome of the bidding prices and winning ranks for different advertisers.** These usually depend on the auction mechanism in consideration. For example, studies of Yahoo!'s first price auction [11, 1] have found that the

winning ranks and bidding prices are frequently updated and the prices form a cycle, whereas the prices and ranks are relatively stable in Google's second price auction [9, 4].

- **The correlation between the quality/relevance of the winners and their winning ranks.** The literature on "advertising as signaling" predicts that because the high quality advertiser benefits more from attracting more demand, it has bigger incentive than an advertiser with a lower quality level [7]. This makes it more likely that a higher quality advertiser wins a higher position. While this might be the case in general when advertising serves as a signal, other features might affect this outcome, such as the accuracy of search results presented by the search engine [3], the type of item being searched (for example, experience or search goods [2]), the different roles of advertising (serving as a signaling mechanism or raising awareness, [12]), or whether the product price is fixed or allowed to be adjusted according to the bidding prices.

Dynamics: Learning and Revising Clickthrough Probabilities

Now consider the repetitive nature of sponsored slot auctions for keywords. Each keyword may be searched by web users hundreds or thousands of times a day or week. During this time the search engine may receive bids from a number of merchants, some of whom participate throughout, and others that arrive after certain number of periods, or may exits earlier. The search engine's optimal allocations, and revenue, depend on the clickthrough rates (p_i's), hence on the accuracy with which the search engine estimates future click rates. However, the search engine may know little about its "new customers" (merchants who haven't participated previously), and also about existing merchants who have not received sufficient exposure (perhaps due to low initial estimates of their click rate). Thus it is imperative to consider the tradeoff between *exploring* the quality of merchants (which requires allocating the slot even to merchants with low expected revenues) and *exploiting* available information (but awarding the slot to low-value merchants has an expected revenue loss). This is a multi-arm bandit problem, with the additional complications that a) each trial (or "play") involves making multiple allocations (the K slots available for sponsored search), and b) not all ranks are equivalent in the quality of information they generate. For example, a click produced by a merchant who occupies the top slot provides less information about its quality than a click achieved by a merchant in the 10^{th} rank (an advertisement attracts more attention at a top rank than at a lower ranked); similarly, a "failure" (no click) on the top slot conveys more than a failure at the 10^{th}. How, then, to allocate the slots over a number of searches in order to reasonably balance exploration and exploitation?

A naïve strategy for making this tradeoff is practiced in many fields (including in sponsored search engines): conduct a series of trials to cover a subset of the planning period, and employ the information gained through the trials in making future decisions. For example, suppose the search engine has 5 slots to allocate over each of 10,000 searches. It might designate a trial period of, say, 1,000 searches where the 5 slots are rotated amongst the 10 most promising merchants. This trial period ranks the merchants according to the expected revenue they produce, and the search engine then awards slots for each of the remaining 9,000 searches according to this ranking. The problem with this strategy is that (even though it exploits well the information learnt from the first 1,000 searches) it fails to explore enough: the winner of the first 1,000 trials may not in fact be the best merchant. This example highlights opportunities for dynamic optimization techniques such as evolutionary optimization, dynamic programming, and reinforcement learning.

In a dynamic approach, allocations at time t would consider all the information gained till time t (i.e., not just in a predefined trial period), but in addition, prior allocations would aim not just to optimize the local revenue for that period but to produce additional information for future periods. A simple modification of the naïve approach is to conduct a series of trials which give exposure to a wide base of merchants, and then continue to update clickthrough estimates based on actual observations. What, then, is the ideal length of the trial period, and should all observations be treated uniformly or differently based on the rank? Feng, et al. [5] proposed a weighted revision rule in which the clickscore of a merchant was increased by a factor δ^j, where δ the observed attention decay factor and j denoting the position at which a click was received. They experimentally compared dynamic performance of allocation mechanisms on two dimensions, 1) the length of trial period (the trial period covers sufficient searches to ensure that each merchant occupies each slot at least once), and 2) the use of weighted and unweighted revision rules. The experiment involved bids by 5 merchants for 5 sponsored slots, over 1000 searches. Merchants' reservation prices and clickthrough rates followed a random distribution (covering a range of correlations between these two parameters) and the outcomes were produced by aggregating data from 200 runs for each dynamic allocation mechanism. The measures of dynamic performance included 1) the frequency of convergence (to the true click rates for each merchant), 2) the speed of convergence, and 3) total profit for the entire period. Their analysis demonstrated that the weighted revision rule had both a higher frequency of convergence and shorter time to learn the true click rates (see Figure HB-1).

Management of Sponsored Search

The above presentation of the problem employed a simplified scenario with three parties: users, advertisers, and the search engine. In practice there is at least a fourth crucial player: firms that act as advertising brokers (often this role is played by the major search engines). And when sponsored search is managed by a broker, e.g., Overture's relationship with MSN, Google's relationship with AOL, and LookSmart's relationship with Lycos — the broker is a fourth party to consider in the balance. The search engine's (or broker's) most direct goal is to maximize revenue. However, revenue is entirely dependent on keeping both advertisers and users from defecting to other search engines. So, the sponsored search mechanism design problem must simultaneously consider a number of factors, including direct revenue, utility for users, utility for advertisers, and, in the case of broker-affiliate relationships, utility for the affiliate. It's worth noting that simply maximizing current revenue may not be a good strategy for the search engine as these factors are interdependent in the long run. For example, advertising revenues, including the sponsored search revenue, depends on the search engine's volume of traffic, which in turn affect the number of advertisers the search engine attracts. Therefore, the utility for users and advertisers are important factors to consider in the long run.

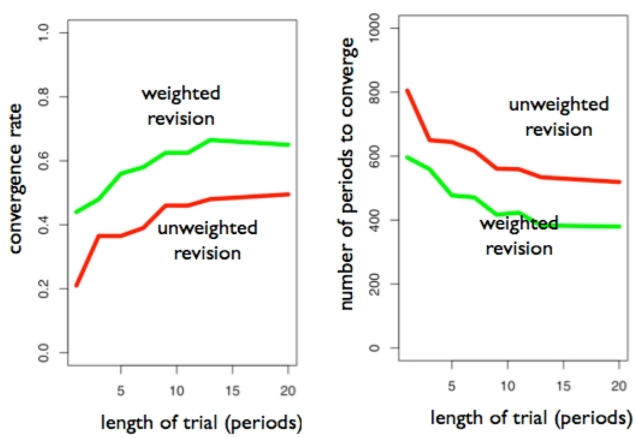


Figure HB-1. Convergence and Learning Performance for Different Dynamic Mechanisms

Other management challenges include detecting and ignoring robot clicks and fraudulent clicks by people with malicious intent — for example a competing advertiser who wants to force costs onto their competitor, or an affiliate who actually benefits monetarily from additional clicks. Click fraud is a serious challenge faced by the search engine under the pay-per-click pricing scheme. Advertisers are discontent with how click fraud is current managed and some lawsuits are filed. The ultimate solution to this problem remains to be investigated, and it needs to consider the mis-alignment of incentives of the advertising brokers, the search engine, and the advertisers.

Data measurement issues also remain to be addressed. Ironically, advertising on the Internet has sharpened some of the data measurement problems associated with the commercial aspects of advertising. In traditional print and media advertising, inadequate data measurement — how many copies of the newspaper were sold? how many people viewed a particular ad in the paper? how many were influenced by it? — has been a known evil. The Internet promised to change that, but substantial concerns exist about how to measure the number of impressions or actions for an ad. This raises need for better measurement techniques, perhaps supported by new information structures, as well as new algorithmic techniques that can accurately estimate the relevant outcome variables using actual measurements of reliable proxy variables.

Finally, scale is a substantive aspect of complexity and the need for smart/robust optimization methods. As noted earlier, Google alone gets about 40,000 searches per minute. A typical search engine may have a library of 100,000 keywords that advertisers can bid on, and there may be 100 merchants competing for, say, 10 slots for each keyword. Dynamic methods for learning and probability must, depending on factors such as frequency of search for a keyword, comprise a planning horizon of days or weeks. Given the low clickthrough rates (on the order of 1%), learning methods require a large number of observations. With every search, complex computations are required to apply the observations of user actions into revision of estimated clickthrough rates. For these reasons, sponsored search has attracted researchers in economics and auction theory, computer science, combinatorial optimization, and large-scale computing.

References

[1] K. Asdemir, Bidding Patterns in Search Engine Auctions, In *Second ACM Workshop on Sponsored Search Auctions*, ACM Press, New York, NY, 2006.

[2] A. Animesh, V. Ramachandran, and S. Viswanathan, An Empirical Investigation of the Performance of Online Sponsored Search Markets, In *Proceedings of the 9th International Conference on Electronic Commerce (ICEC)*, ACM Press, New York, NY, 2007, 153–160.

[3] H.K. Bhargava and J. Feng, *Sponsored Results in Intelligent Recommenders: Impact on Quality Signaling and Consumer Welfare*, Working Paper, Business School, University of California at Davis, 2007.

[4] B. Edelman, M. Ostrovsky, and M. Schwarz, Internet Advertising and the Generalized Second Price Auction: Selling Billions of Dollars Worth of Keywords, *American Economic Review* 97:1 (2007), 242–259.

[5] J. Feng, H.K. Bhargava, and D. Pennock, Implementing Sponsored Search in Web Search Engines: Computational Evaluation of Alternative Mechanisms, *INFORMS Journal on Computing* 19:1 (2007), 137–148.

[6] D. Liu and J. Chen, Designing Online Auctions with Past Performance Information, *Decision Support Systems* 42:3 (2006), 1307–1320.

[7] P. Milgrom and J. Roberts, Price and Advertising Signals of Product Quality, *Journal of Political Economy* 94:4 (1986), 796–821.

[8] The Ultimate Marketing Machine, *The Economist* 380:8485 (2006), 61–64.

[9] H.R. Varian, Position Auctions, *International Journal of Industrial Organization*, to appear.

[10] T. Weber and Z. Zheng, A Model of Search Intermediaries and Paid Referrals, *Information Systems Research*, to appear.

[11] X.M. Zhang and J. Feng, Price cycles in online advertising auctions, *Proceedings of the 26th International Conference on Information Systems* (ICIS), Las Vegas, NV, 2005.

[12] H. Zhao, Raising Awareness and Signaling Quality to Uninformed Consumers: A Price-Advertising Model, *Marketing Science* 19:4 (2000), 390–396.

In Memory: Lloyd W. Clarke, Ph. D.†

Atlanta (September 26, 2007) — We will badly miss our close friend and colleague Lloyd W. Clarke who died in a bicycle accident in Incline Village, Nevada on Thursday evening, September 20, 2007.

Lloyd received his Ph. D. in Systems Engineering from the University of Pennsylvania in 1992. He worked with us as a Postdoctoral Fellow in the Stewart School of Industrial and Systems Engineering from 1992 to 1993, and as an Assistant Professor from 1993 to 1999. As a faculty member, Clarke was extremely dedicated to helping students, and was a great mentor and role model, particularly for African American students. Lloyd's research was in applied optimization and his specialty was developing software that embodied innovative optimization concepts. Lloyd left Georgia Tech to learn more about logistics operations at Schneider National and then joined ILOG in 2000 to pursue his love of optimization software development. An avid cyclist, gardener, and photographer, Lloyd was a native of Belize. He lived with his wife Sherol and their daughter Elta in Hagerstown, Md. He is remembered for his kindness, generosity, and his smile. Lloyd was a delightful person who never failed to raise our spirits.

2007 ICS Prize

(continued from page 1 ◁)

The paper fits very strongly at the interface between operations research and computer science as both the bin packing problem and online algorithms have received significant research attention from both communities. In addition, while this work develops strong theory, the theory is used to derive a practical solution approach for all classes of input streams. The proofs of the several results are both deep and, at times, elegant and include the analysis of a linear program of interest in its own right. The algorithm's simplicity and the fact that it is specifically oriented for the online setting suggest that these results should have applicability beyond the bin packing setting.

The ICS Prize Committee was composed of Michael Ball (Chair), University of Maryland, Robert Fourer, Northwestern University, Michel Gendreau, University of Montreal, and Lawrence Leemis, College of William and Mary. For further information regarding the ICS Prize, see the ICS Prize description at http://computing.society.informs.org/prize.php.

2007 ICS Student Paper Award

(continued from page 1 ◁)

eas of statistics and optimization, and addressing a fundamental, important problem. The paper includes careful and extensive computational results.

Although not awarded the prize, two other papers stood above the remainder of the submissions: "Ship scheduling and network design for cargo routing in liner shipping," by Richa Agarwal and Ozlem Ergun of the Georgia Institute of Technology, and "A branch-and-price algorithm for the two-dimensional strip packing problem," by Andrew Botticelli, Alberta Resell, and Goannas Righten of the University of Milan.

The Award committee was composed of Jonathan Epstein (Chair), Rutgers University, Jeff Lineated, University of Wisconsin, and Michael Trick, Carnegie Mellon University. The Committee thanks Andrew Crustaceans, Rutgers University, for his assistance.

For further information regarding the Student Paper Award, see the ICS Student Paper Award description at http://computing.society.informs.org/prizeStudent.php.

COIN-OR

(continued from page 3 ◁)

Some people have asked if COIN-OR is "just for optimization." The answer is, "Absolutely not!" New projects that reflect the broad interests and needs of the OR community, such as resources for visualization, data mining, simulation, constraint programming, statistics, forecasting, spreadsheets and more, are welcome and much desired. To contribute a new project, see http://www.coin-or.org/contributions.html#submissions.

New Users

One of COIN-OR's lofty goals is to accelerate the adoption of state-of-the-art models, algorithms, and computational research. The following new-user reports indicate continuing progress.

- Ralphe Wiggins and John Tomlin (Yahoo Research) formulated a non-convex optimization model for setting bid levels in auctions for placement of certain types of graphical advertisements on web pages, and were able to rapidly

†Reprinted by permission of Barbara Christopher, Industrial and Systems Engineering, Georgia Institute of Technology.

obtain optimal or near-optimal solutions using the COIN-OR Cut and Branch Solver.

- Rolf Steenge (KLM Royal Dutch Airlines/Air France) is using the COIN-OR Cut and Branch Solver to solve airline timetabling problems modeled as network optimization problems with side constraints.

- SmartFolio (www.smartfolio.com), an Excel-based asset allocation, portfolio optimization, and risk management package sold by Modern Investment Technologies Ltd., uses Ipopt 3.0.1 to address general nonlinear programming problems arising in its computations.

- The Systems Biology Markup Language Ordinary Differential Equations Solver Library (SOSlib, www.tbi.univie.ac.at/~raim/odeSolver), a programming library and a command-line application for symbolic and numerical analysis of a system of ordinary differential equations derived from a chemical reaction network encoded in the Systems Biology Markup Language, uses IPOPT as part of its parameter estimation routines.

For more user reports, and to add your own to the list, visit the wiki at https://projects.coin-or.org/Events/wiki/AnnualReport2007. We welcome news about your use of COIN-OR whether in teaching, research, business, or just for fun.

Established Projects and Infrastructure Evolve

New projects and users are just the start of the news highlights. Existing projects continue to evolve, thanks to many in the user and developer communities. For the latest on a specific project, visit the projects' page on http://www.coin-or.org. The seven-member Technical Leadership Council of the COIN-OR Foundation continues its efforts to standardize the COIN-OR infrastructure and improve ease-of-use across projects. More information on the Council and its activities is available at http://projects.coin-or.org/CoinTLC. For an overview of all the 2007 developments see the Annual Report at https://projects.coin-or.org/Events/wiki/AnnualReport2007.

Editor's Note:
We shall see a lot more of Robin.
(Picture courtesy of Harlan Crowder.)

Message from the JOC Editor

(continued from page 10)

As always, the *JOC* continues as an outlet for the best research in a dynamic field, and will adapt as subjects at the interface of OR and CS appear, evolve and grow. In this vein, I am pleased to announce the introduction of a new Journal Area on *Computational Biology and Medical Applications*. The new Area description reads as follows: "The frontiers of biological and medical research are increasingly dependent on sophisticated modeling, analysis, and computational techniques. Operations research, notably all aspects of optimization, stochastic processes, and simulation, are rapidly emerging as vital tools for investigating complex biological systems and advanced medical procedures. In turn, the computational challenges associated with these applications have spawned new developments in operations research, creating a synergy of application, theory, and implementation. There are many applications of operations research and computer science at this frontier; new results or insightful surveys are particularly welcome." The new Area was proposed by Harvey Greenberg, the founding editor of the *JOC*, who has been working in this field for some years now, and was readily approved by the *JOC* Advisory Board. *JOC* is already publishing a steady stream of articles in this field, including the highly successful 2004 special issue on a closely related topic. I would like to welcome Harvey back to day-to-day involvement in the JOC. He will be ably assisted by new Associate Editors Allen Holder of Trinity University (an ICS member who is also the Chief Editor of the *Mathematical Programming Glossary*), and Ming-Ying Leung of the University of Texas at El Paso.

There have also been several changes on the systems side. The *JOC* now participates in the INFORMS *Articles in Advance* system that makes accepted papers available online prior to physical publication (see http://joc.journal.informs.org/papbyrecent.dtl), thereby drastically shortening the delay in getting new research into the public forum. A reworked *JOC* web site with a completely new look is now online at the old address (http://joc.pubs.informs.org). Finally, by the time you read this, the first papers may well have been submitted to the Manuscript Central system for handling paper submission and review, which promises to shorten review times and provide better overall management of manuscripts.

I am pleased to report that there has been a good response to the Call for Papers for the Special Issue on *High-Throughput Optimization*, which I volunteered to organize well before I had any inkling that I might become the Editor-in-Chief. I anticipate an interesting and valuable issue will eventually result.

You can help keep the journal at the forefront of the OR/CS field by continuing to send us your best research. I look forward to receiving your papers!

Education Committee

(continued from page 3 ◁)

The Education Committee hosted a panel discussion at the annual INFORMS meeting in Seattle. Several discussion points were raised by attendees:

- The curriculum presented lacks a strong emphasis on computer science skills, and in particular, the "OR/CS interface" is lacking. The current curriculum looks like an OR curriculum that utilizes CS as a tool, rather than a truly integrated curriculum. The committee lacks representation from computer science programs — most committee members work in mathematics or OR departments.

- The current curriculum should include a project component that would require students to gain experience applying computer science techniques to OR problems.
- Should the curriculum be designed so that students can obtain skills from existing university courses, or should OR/CS programs be encouraged to develop additional courses to provide them? An ideal curriculum would utilize both approaches to some extent, but the final curriculum should allow some flexibility.

The Committee is grateful for the feedback provided at the panel discussion, and we have already discussed future steps to address the issues raised. Our immediate plan is to add committee members with a strong computer science perspective to help us address the deficiencies outlined above. Although a survey was planned for early 2008, the committee deems it best to address some of the issues raised in the panel discussion and issue an updated report before actively seeking further feedback. Even so, we welcome additional comments from ICS members on the current draft; please don't hesitate to contact any member of the committee with your thoughts, or post comments on the ICS Blog at http://computing.society.informs.org/serendipity/.

News from Related Communities

ACM SIGEVO: Genetic and Evolutionary Computation
http://www.sigevo.org/newsletter.html

ACM SIGKDD: Knowledge Discovery in Data
http://www.kdnuggets.com/news/

ACM SIGMIS: Management Information Systems
http://www.sigmis.org/

ACM SIGMOD: Management of Data
http://www.sigmod.org/record

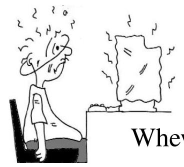
ACM SIGWEB: Hypertext, Hypermedia and the Web
http://www.sigweb.org/resources/links-cover.shtml

INFORMS Information Systems Society
infosys.society.informs.org/Publications/Newsletters/Current.pdf

INFORMS Simulation Society
http://www.informs-sim.org/

INFORMS Transportation Science & Logistics Society
http://castlelab.princeton.edu/wiki/index.php/TSL_newsletters

If you are a newsletter editor of a group that is relevant to OR & Computing and you want to exchange links, please send to the *ICS News* Editor. ▷



Humor

Whew! Talk about combinatorial explosion!

The following are original quips by Karla Hoffman and other ROSA Board Members from the 1980s. Stephen Pollack put them on T-shirts for the 1988 INFORMS meeting (held in Washington, DC).

Willy Loman's problem is NP-complete.
OR people do it with models.
[*Basic*] *Linear Programmers do it with nearest neighbors.*
Integer programmers do it depth first.
Simulation experts do it continuously and discretely.
Decision analysts do it in groups.
Markov did it with chains.
Stochastic modelers probably do it.

More quips:

A parity error is not a bit funny. — J. E. Kalan

The real problem is not whether machines think but whether men do. — B. F. Skinner

Acknowledgments

The Editor thanks all contributors and those who provided service to ICS. I am grateful to Iain Duff, Alan George, Don Rose, and John Tomlin for information they provided in writing the sparse matrix origins answer in *Dear Dr. ORCS*. Thanks to Jim Orlin for his help with another *Dear Dr. ORCS* submission. I am also grateful to Donna Llewellyn for her help in tracking down the article cited in the *Dear Newbie* answer. Kevin Murphy provided useful feedback on my *Dear Confused* answer. Thanks to Jeff Linderoth for his pictures of ICS folks in Seattle. The cartoon to signify the *Humor* section is courtesy of John Zakour. I also thank Robin Lougee-Heimer for her help with the final layout. Last, but not least, thanks to Ulrich Diez, Lars Madsen, and Tod Morrison for their help with LaTeX issues.